U12-type Spliceosome: Localization and Effects of Splicing Efficiency on Gene Expression

Heli Pessa

Institute of Biotechnology
and
Division of Genetics
Department of Biosciences
Faculty of Biological and Environmental Sciences
and
Viikki Graduate School in Biosciences

University of Helsinki

Academic dissertation

To be presented for public examination with the permission of the Faculty of Biological and Environmental Sciences of the University of Helsinki in Auditorium 1041 in Biocenter 2, Viikinkaari 5, Helsinki, on 15 October, 2010 at 12 o'clock noon.

Supervisor Docent Mikko Frilander
Institute of Biotechnology
University of Helsinki
Helsinki, Finland

Reviewers Docent Marjo Salminen
Department of Basic Veterinary Sciences
Faculty of Veterinary Medicine
University of Helsinki
Helsinki, Finland

Professor Ykä Helariutta
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

Opponent Professor Marie Öhman
Department of Molecular Biology and Functional Genomics
Stockholm University
Stockholm, Sweden

Custos Professor Tapio Palva
Department of Biosciences
Faculty of Biological and Environmental Sciences
University of Helsinki
Helsinki, Finland

ISSN 1795-7079
ISBN 978-952-10-6438-8 (paperback)
ISBN 978-952-10-6439-5 (PDF, http://ethesis.helsinki.fi)

Cover image: U12 snRNA detected by fluorescent *in situ* hybridization in mouse NIH-3T3 cell nuclei
Cover layout by Timo Päivärinta

Helsinki University Print, Helsinki 2010

Scientists have one thing in common with children: curiosity. To be a good scientist you must have kept this trait of childhood, and perhaps it is not easy to retain just one trait.

– Otto Robert Frisch

Contents

List of original publications

This thesis is based on the following articles, which are referred to in the text by their Roman numerals. The articles are reproduced with kind permission from the copyright owners.

I Pessa, H.K.J., Ruokolainen, A. and Frilander, M.J. (2006) The abundance of the spliceosomal snRNPs is not limiting the splicing of U12-type introns. RNA *12*: 1883-1892.

I Pessa, H.K.J., Will, C.L., Meng, X., Schneider, C., Watkins, N.J., Perälä, N., Nymark, M., Turunen, J.J., Lührmann, R. and Frilander, M.J. (2008) Minor spliceosome components are predominantly localized in the nucleus. Proc. Natl. Acad. Sci. U.S.A. *105*: 8655-8660.

II Pessa, H.K.J., Greco, D., Kvist, J., Wahlström, G., Heino, T.I., Auvinen, P. and Frilander, M.J. Gene expression profiling of U12-type spliceosome mutant *Drosophila* reveals widespread changes in metabolic pathways. PloS ONE, in press.

Author's contribution to the publications

I HP performed experiments and participated in data analysis and writing of the manuscript.

II HP designed and performed part of the experiments, analyzed data and wrote the manuscript.

III HP designed and performed experiments, analyzed data and wrote the manuscript.

Abbreviations

A	Adenine
ASF	Alternative splicing factor
ATP	Adenosine triphosphate
C	Cytosine
EJC	Exon junction complex
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
G	Guanine
GO	Gene ontology
GTP	Guanosine triphosphate
hnRNP	Heterogenous nuclear RNP
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
mRNA	Messenger RNA
nt	Nucleotide
PCR	Polymerase chain reaction
PHB	Prohibitin
qRT-PCR	Quantitative RT-PCR
R	Purine (A or G)
RNA	Ribonucleic acid
RNP	Ribonucleoprotein
RS	Arginine/serine-rich (domain)
RT-PCR	Reverse transcription PCR
SF	Splicing factor
SMA	Spinal muscular atrophy
SMN	Survival of motor neurons
snRNA	Small nuclear RNA
snRNP	Small nuclear RNP
SR	Serine/arginine-rich (protein)
T	Thymine
tRNA	Transfer RNA
U	Uracil
U2AF	U2 auxiliary factor

Abstract

The removal of non-coding sequences, introns, is an essential part of messenger RNA processing. In most metazoan organisms, the U12-type spliceosome processes a subset of introns containing highly conserved recognition sequences. U12-type introns constitute less than 0,5% of all introns and reside preferentially in genes related to information processing functions, as opposed to genes encoding for metabolic enzymes. It has previously been shown that the excision of U12-type introns is inefficient compared to that of U2-type introns, supporting the model that these introns could provide a rate-limiting control for gene expression.

In this work, cells with low abundance of the U12-type spliceosome were found to inefficiently process U12-type introns encoded by a transfected construct, but with endogenous genes, the abundance of the U12-type spliceosome was not found to affect expression levels. However, significant levels of endogenous unspliced U12-type intron-containing pre-mRNAs were detected in cells. Together these results support the idea that U12-type splicing may limit gene expression in some situations.

The effect of U12-type splicing efficiency on a whole organism was studied in a *Drosophila* mutant deficient in U12-type splicing. Genes containing U12-type introns showed variable gene-specific responses to the splicing defect. Surprisingly, microarray screening revealed that metabolic genes were enriched among downstream effects, and that the phenotype could largely be attributed to one U12-type intron-containing mitochondrial gene. Gene expression control by the U12-type spliceosome could thus have widespread effects on metabolic functions in the organism.

The subcellular localization of the U12-type spliceosome components was studied as a response to a recent dispute on the localization of the U12-type spliceosome. All components studied were found to be nuclear indicating that the processing of U12-type introns occurs within the nucleus, thus clarifying a question central to the field.

1 Review of the literature

1.1 Overview of eukaryotic pre-mRNA processing

A eukaryotic gene typically consists of several stretches of coding sequence, exons, separated by non-coding introns (Figure 1). The discontinuous structure of eukaryotic genes presents a problem for the gene expression machinery: open reading frames separated by intronic areas must be joined together to produce a functional messenger RNA (mRNA). This process, mRNA splicing, is achieved by the spliceosome, a large complex of small nuclear RNA (snRNA) molecules assembled into small nuclear ribonucleoprotein (snRNP) particles and numerous auxiliary proteins. Intron recognition and spliceosome assembly are followed by catalytic reactions that result in the joining of exons and release of the intron. Splicing occurs primarily cotranscriptionally, and is integrated with other pre-mRNA processing events in the nucleus, such as capping and polyadenylation. Transcripts become mature mRNAs through multiple processing steps prior to transport out of the nucleus into the cytoplasm, where they can be translated, stored or degraded.

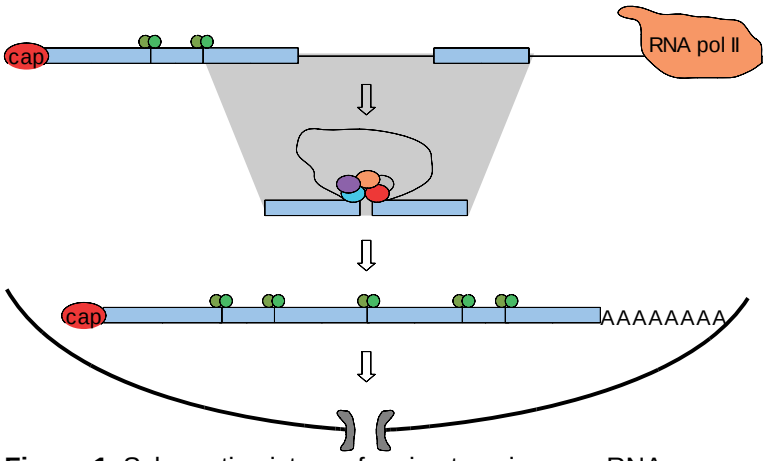

Figure 1. Schematic picture of main steps in pre-mRNA processing. Blue boxes, exons; thin black lines, introns; green dots, exon junction complex; thick line, nuclear membrane and pore.

1.2 Introns in eukaryotic genes

1.2.1 Complex organisms have large variation of intron sizes

In higher eukaryotes, multiple introns are present in most genes. Mammals in particular have a surprising variety of intron sizes and numbers per gene. In humans, mean intron length is 3300 bp, whereas in the fruit fly *Drosophila melanogaster*, the average length is less than 500 bp (Lander *et al.* 2001). An extreme example is provided by the longest human gene, coding for the muscle-specific protein dystrophin. It spans 2,5 million bases, but its 79 exons account less than 1% of total length, with the average intron length of 26 000 bases (Pozzoli *et al.* 2002). In general, simple eukaryotes have less introns than multicellular organisms, but the frequency of introns varies in different lineages (reviewed by Jeffares *et al.* 2006). In baker's yeast *Saccharomyces cerevisiae*, only 4% of genes have introns, usually only one per gene (Kupfer *et al.* 2004). Intron size distribution in yeast, from 50 to 1000 nt, is more narrow than in more complex organisms.

1.2.2 Introns early or introns late

The origin of the discontinuous structure of eukaryotic genes is unclear. Two alternative scenarios have been suggested for the introduction of introns into eukaryotic genomes: the 'introns early' and 'introns late' hypotheses (Darnell 1978, Doolittle 1978; for a review, see e. g. Koonin 2006). According to the former theory, reading frames have never been continuous, and prokaryotes have lost their introns as their genomes have become more compact. The 'introns late' theory depicts that introns have been inserted into originally continuous genes throughout eukaryotic evolution. The 'introns early' model, included in the original version of the 'exon theory' of gene evolution formulated by Gilbert (1987), assumes that in the earliest living organisms with protein-coding genes, each exon was a separate gene coding for a distinct protein domain, which allowed proteins to be functional even before the existence of splicing mechanisms. The development of splicing then led to the formation of longer genes with several exons. As a consequence, present-day genomes should contain an excess of phase 0 introns, which are located between codons rather than inserted within a codon. Phase 0 introns are indeed more common in many genomes than phase 1 or 2 introns, but other explanations have been presented, such as the preferential fixation of phase 0 introns (Sverdlov *et al.* 2003). While some reports have found a correlation between protein domain boundaries and intron positions, the theory has not received convincing support (Stoltzfus *et al.* 1994, de Souza *et al.* 1996).

1.2.3 Introns are gained and lost

According to a widely accepted synthesis between different theories on the origin of introns, intron gain has been extensive in the common eukaryotic ancestor, but loss and gain of introns have continued later, at different rates in different lineages (Roy and Gilbert 2005). Ancestral introns are suggested to have arisen as the result of the insertion of mobile elements, such as transposons (Crick 1979, Sharp 1985). Reversibility of the excision of self-splicing introns has promoted the hypothesis that an excised intron could reinsert into another mRNA and subsequently be reverse transcribed and recombined into the genome by homologous recombination (Kruger *et al.* 1982). Reverse splicing of a nuclear intron by the spliceosome has indeed been demonstrated *in vitro* (Tseng and Cheng 2008). Introns can be lost through homologous recombination with a spliced and reverse-transcribed mRNA or deletion via nonhomologous recombination (Robertson 1998). The former process is believed to have lead to the loss of most introns in yeast, causing a biased distribution of the remaining ones. Since reverse transcription begins at the 3' end, introns close to the 3' end have a greater chance of being lost, which explains the excess of old introns near the 5' ends of genes in intron-poor genomes (Sverdlov *et al.* 2004). The introduction of introns into the ancestor of eukaryotes has been speculated to have contributed to the development of the nucleus (Martin and Koonin 2006). In this model, the endosymbiont that gave rise to mitochondria infected its host with self-splicing introns, which proliferated extensively in the host genome. The nuclear membrane was then needed to isolate unspliced pre-mRNAs from the translation machinery until splicing had been completed. However, the order of appearance of eukaryote-specific traits has not been resolved.

The removal of introns from pre-mRNAs requires a great deal of resources, but introns also benefit the organism. Alternative splicing is an important way of generating transcript

variation especially in complex eukaryotes (see **1.5**). It has also been suggested, in the 'exon shuffling' hypothesis, that the discontinuous structure of genes accelerates the evolution of new genes by combining exons in new ways (Tonegawa *et al.* 1978, Gilbert 1978, Vibranovski *et al.* 2005). The advantage of discontinuous genes in this model is that recombination inside introns need not be precise to avoid disruption of the coding sequence.

1.3 Intron types

1.3.1 Two types of spliceosomal introns in eukaryotes

Most introns are processed by the canonical U2-type spliceosome. They contain only short stretches of highly conserved sequence; at the splice sites, only the terminal dinucleotides GT-AG are 99% conserved. The polypyrimidine tract near the 3' splice site and the branch point sequence, usually 20-40 nt upstream of the 3' splice site and containing the branch point nucleotide, almost always an adenosine, show some conservation (Burge *et al.* 1998, Sheth *et al.* 2006). Yeast introns have longer conserved sequences at the splice sites and a weak polypyrimidine tract (Rymond and Rosbash 1985, Fouser and Friesen 1987).

In the early 1990s, a subset of introns with long and nearly invariant consensus sequences were discovered (Jackson 1991). Based on their complementarity with the low-abundance snRNAs U11 and U12, these introns were suggested to depend on a different set of snRNAs for their splicing than the common intron type (Hall and Padgett 1994). The hypothesis was later confirmed by mutational and biochemical analysis (Hall and Padgett 1996, Tarn and Steitz 1996a, Tarn and Steitz 1996b). The first minor class introns found had terminal AT-AC dinucleotides, hence the minor splicing system was originally dubbed AT-AC splicing. However, many minor class introns have conventional GT-AG termini (see **1.3.2**), and AT-AC introns were renamed U12-dependent after one of the snRNAs required for their processing (Dietrich *et al.* 1997, Sharp and Burge 1997). U12-dependent, or U12-type, introns have been found in most multicellular organisms with the nematode *Caenorrhabditis elegans* being the most prominent exception (Burge *et al.* 1998, Levine and Durbin 2001, Zhu and Brendel 2003, Sheth *et al.* 2006, Alioto 2007). Genomic information from an increasing number of species has allowed the identification of U12-type introns and spliceosomal components also in many simple eukaryotes, including the nematode *Trichinella spiralis*, protists *Acanthamoeba castellanii* and *Physarum polycephalum*, several fungal lineages and three species of genus *Phytophthora* (Russell *et al.* 2006, Dávila López *et al.* 2008). Many other unicellular organisms, e.g. the yeasts *S. cerevisiae* and *Schizosaccharomyces pombe* are lacking the U12-type system. Computational analyses on intron frequencies have revealed a higher proportion of U12-type introns in more complex eukaryotes. In human and mouse, U12-type introns make up about 0,36% of all introns whereas in the plant A*rabidopsis thaliana*, the proportion of U12-type introns is 0,24%. Drosophila has only about 20 U12-type introns (Schneider *et al.* 2004, Sheth *et al.* 2006, Lin *et al.* 2010). The scarcity of U12-type introns is not typical of all insects but is characteristic of the dipteran group, in which the U12-type spliceosome has apparently undergone a rapid evolution (Mount *et al.* 2007, Lin *et al.* 2010).

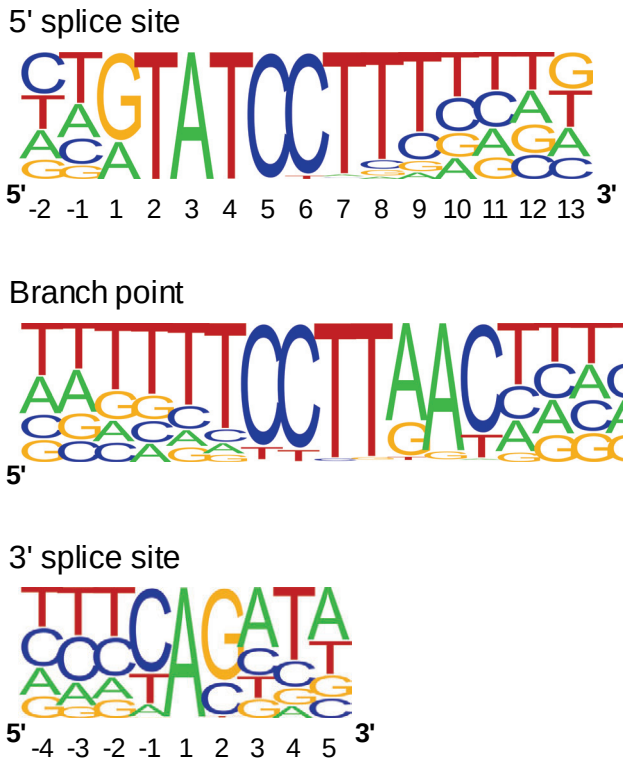

Figure 2. Sequence logos of human U12-type 5' splice site, branch point and 3' splice site. The height of the letters indicates the frequency of bases at each position. Numbering is relative to splice junctions. Frequency logos were created with the Weblogo program (Crooks et al. 2004, http://weblogo.berkeley.edu) using sequences downloaded from the U12 database (Alioto 2007).

1.3.2 U12-type consensus sequences

The terminal dinucleotides do not distinguish between major and minor introns, because many U12-type introns have conventional GT-AG termini, and a few introns with AT-AC termini are spliced by the U2-type spliceosome (Dietrich *et al.* 1997, Wu and Krainer 1997, Sharp and Burge 1997). In humans, 70% of U12-type introns are of the GT-AG subtype, 25% belong to the AT-AC subtype, and the rest have non-canonical terminal dinucleotides (Sheth *et al.* 2006). U12-type introns have longer conserved sequences at the 5' splice site and branch point than U2-type introns (Figure 2). U12-type 5' splice site has the consensus sequence RTATCCTT, where R is a purine. U12-type introns lack a polypyridimine tract, but the branch site 10-20 nucleotides upstream of the 3' splice site has the consensus sequence TCCTTAAC, although the conservation is less rigid than at the 5' splice site (Figure 2; Hall and Padgett 1994, Brock *et al.* 2008). A small number of mutations at the 5' splice site and branch point may cause the switching of a U12-type intron into U2-type, but the reverse process is highly unlikely due to the extent of conserved sequences required by the U12-dependent spliceosome (Burge *et al.* 1998). There is, however, one documented case of U12-type intron gain in dipteran insects (Lin *et al.* 2010). The tendency of U12-type introns to be converted into U2-type introns in evolution can be deduced from several cases in which a normal intron is present in a homologous location in a related species. *C. elegans* is believed to have lost U12-type introns and the U12-dependent spliceosome from its genome through this conversion process. However, the locations of U12-type introns are more conserved between species than the positions of U2-type introns (Basu *et al.* 2008). Some U12-type intron locations are remarkably conserved between organisms as distant as human and *Arabidopsis* (Zhu and Brendel 2003). This amazing conservation over evolutionary distance has promoted the hypothesis that U12-type introns have an important function in gene expression.

1.3.3 U12-type introns may have a regulatory function

The reason for the persistent retention of U12-type introns in most metazoan genomes has been hypothesized to be their possible function in expression level control. U12-type

introns are spliced less efficiently than U2-type introns, as indicated by the presence of pre-mRNAs with unspliced U12-type introns in cells (Patel *et al.* 2002). In the report by Patel *et al.* (2002), conversion of a U12-type intron into U2-type led to a 6- to 8-fold increase in the expression levels of the reporter protein. The distribution of U12-type introns in the genome is not entirely random, but minor introns tend to reside in genes that function in 'information processing' tasks, such as DNA replication or repair, transcription, RNA processing and translation, rather than in genes coding for metabolic enzymes. In addition, the number of genes containing two U12-type introns is larger than would be expected if the locations of U12-type introns were random (Burge *et al.* 1998; reviewed by Wu and Krainer 1999).

1.3.4 Non-spliceosomal introns

In addition to spliceosomal introns, many organisms contain self-splicing introns (Kruger *et al.* 1982). They are found in bacteria and in the organelles of plants, fungi and protists (see review by Bonen and Vogel 2001). Self-splicing introns are divided in groups I and II according to their splicing mechanism. Group II introns are excised through reactions that resemble the splicing of introns by the spliceosome: a two-step transesterification reaction with adenosine as the initiating nucleophile, producing a lariat intron. Self-splicing introns can catalyze their own removal *in vitro*, but protein factors are required *in vivo*. The intron often contains an open reading frame that encodes for functions required for splicing and mobility, such as endonuclease and reverse transcriptase, but factors produced by the host genome are also needed. Self-splicing introns apparently have a common evolutionary history with retrotransposons.

Some tRNA genes contain introns that are spliced through a mechanism different from spliceosomal or self-splicing reactions. The splicing of tRNA introns involves an endonuclease, an RNA ligase and a phosphotransferase, and requires energy from ATP hydrolysis (reviewed by Abelson *et al.* 1998).

1.4 Splicing signals

1.4.1 Exon definition facilitates intron recognition

Most human exons are relatively short, with length distribution peaking around 100 bp (Lander *et al.* 2001), but introns can be several kilobases long and contain sequences resembling splice sites by chance. Splicing over long introns is facilitated by pairing splice sites across exons instead. In this 'exon definition' model, recognition of the 5' splice site by U1 snRNP promotes the formation of a spliceosomal complex upstream of the flanking exon (Robberson *et al.* 1990). Consistent with the model, vertebrate exons are usually less than 300 bases long, much shorter than introns, and increasing the length of an exon inhibits the removal of the preceding intron. Furthermore, mutations in a downstream 5' splice site inhibit splicing at the upstream intron and cause skipping of the exon between (Talerico and Berget 1990). Exon definition interactions are mediated by U1 snRNP, which enhances the binding of the U2 auxiliary factor 65K subunit (U2AF-65) at the branch point of the upstream intron (Hoffman and Grabowski 1992). Splicing of the first and last exon is facilitated by the 5' cap binding complex and the polyadenylation machinery, respectively, through interactions involving U1 snRNP (Niwa and Berget 1991, Wassarman and Steitz 1993, Lewis *et al.* 1996). Exon bridging interactions have been detected also between U2- and U12-type spliceosomes

(Wu and Krainer 1996). In Drosophila, the problem of large introns may be overcome by recursive splicing at elements that first function as 3' splice sites and then generate 5' splice sites after ligation with an upstream exon (Hatton *et al.* 1998, Burnette *et al.* 2005). Short Drosophila introns are often removed through 'intron definition', in which the splice sites are paired across the intron. This idea is supported by findings that in the fly, 5' splice site mutations cause intron inclusion rather than upstream exon skipping, and increasing the size of an intron inhibits splicing (Talerico and Berget 1994). Further support to differences in intron recognition between evolutionary lineages is provided by a computational analysis by Lim and Burge (2001), according to which 5' and 3' splice site signals in *D. melanogaster* and *C. elegans* contain enough information for the identification of most short introns. In human or *Arabidopsis thaliana*, the information content of splice sites is lower, consistent with the use of exon definition mechanism. The conserved 5' splice site sequence in yeast can accurately identify an intron but, due to the short 3' splice site consensus, exact recognition of the 3' end of the intron requires additional signals. In *S. cerevisiae*, homologues to mammalian factors responsible for exon-spanning interactions are mostly missing (Abovich and Rosbash 1997). Also in the fission yeast *S. pombe*, splicing probably occurs exclusively by intron definition (Romfo *et al.* 2000).

1.4.2 Exonic and intronic sequences influence splice site choice

Splicing fidelity depends on the accurate recognition of introns. Especially in complex eukaryotes in which alternative splicing is common, splice sites are too weak and degenerate to accurately identify intron boundaries without other signals. Human 5' splice site-like sequences occur in average about every 300 bases in random sequence (Burge *et al.* 1999). Since introns are often much longer, the information content in U2-type splice sites is not enough to accurately specify intron ends. The long consensus sequences of U12-type 5' splice site and branch point provide more information for the recognition of splice sites but, due to the rarity of U12-type introns, the consensus sequences are still insufficient for locating the introns. Additional signals are provided by exonic splicing enhancers (ESEs) and silencers (ESSs) within exons and intronic splicing enhancers (ISEs) and silencers (ISSs) in flanking introns (reviewed by Wang and Burge 2008). These elements function as binding sites for factors that either recruit spliceosome subunits to the adjacent splice sites or antagonize spliceosome assembly. Multiple ESEs are commonly found in both constitutive and alternatively spliced human exons (Fairbrother *et al.* 2002). They are usually bound by members of the serine/arginine-rich (SR) protein family, which facilitate spliceosome formation by protein interactions (see **1.5.2**). Also the U12-type spliceosome can respond to ESEs, and SR proteins are required for the splicing of U12-type introns (Wu and Krainer 1998, Dietrich *et al.* 2001, Hastings and Krainer 2001). ESSs are typically recognized by members of the heterogenous nuclear RNP (hnRNP) protein class, which inhibit splicing by various mechanisms. Intronic splicing regulation elements are less well characterized, but they include binding sites for hnRNP proteins and for some tissue-specific splicing factors. Intronic elements play an important role in alternative splicing (Sorek and Ast 2003). The splicing of some U12-type introns is stimulated by the binding of the hnRNP H protein to intronic G-rich sequences (McNally *et al.* 2006).

1.5 Alternative splicing

1.5.1 Alternative splicing increases variety in the transcriptome

Many genes have alternative splicing patterns that result in the production of a variety of mRNA isoforms from one gene (reviewed by Graveley 2001, Wang and Burge 2008). As a general tendency, in simple organisms only a small proportion of genes are alternatively spliced, whereas in more complex eukaryotes, alternative splicing is rather a rule than exception. More than 90% of human genes are alternatively spliced (Wang *et al.* 2008). Isoform patterns take many forms, some of which are depicted in Figure 3. Alternative splicing events include alternative 5' and 3' splice sites and cassette exons that are included or skipped in a regulated manner, such as in the mRNA coding for tropomyosine related kinase B, a neurotrophin receptor (Luberg *et al.* 2010). Further variation can be produced with alternative exons, one of which will be included in the mRNA, illustrated by the *Drosophila* Dscam (Down syndrome cell adhesion molecule) gene. Its mRNA contains several alternative exons, each of which are chosen from an array of mutually exclusive exons (Schmucker *et al.* 2000). Regulated intron retention can function as a means of gene expression control (Mansilla *et al.* 2005). Alternative patterns can be e. g. cell- or tissue-specific (Yeo *et al.* 2004), developmentally regulated, such as the twintron in *Drosophila prospero* gene (see **1.5.3**), or changed according to environmental cues or internal signals, such as the alternative splicing of the protein kinase C mRNA in response to insulin signalling (Chalfant *et al.* 1998). Alternative isoforms may differ in function, their specificity can be altered by a change of a few nucleotides, or entire protein domains may be skipped or introduced. The inclusion or exclusion of regulatory or localization domains may thoroughly change the function of the protein, as in the production of soluble versus membrane-anchored variants of the Fas receptor that have opposing effects on apoptosis (Cascino *et al.* 1995). Alternative splicing does not only induce changes in the coding information, but it may also serve as a regulatory point in gene expression through changes in sequences that affect mRNA stability. In some cases, gene expression can be regulated by the production of splicing isoforms with a premature stop codon that targets the transcript for the nonsense-mediated decay pathway (Lewis *et al.* 2003; see also **1.10.2**). Examples of such level regulation are the alternative splicing events on the mRNAs of U11/U12-48K and 65K proteins. The former case involves the creation of a premature stop codon, whereas the latter produces two mRNA isoforms with different half-lives (Verbeeren *et al.* 2010). The levels of several other core spliceosomal proteins are similarly controlled by nonsense-mediated decay (Saltzman *et al.* 2008).

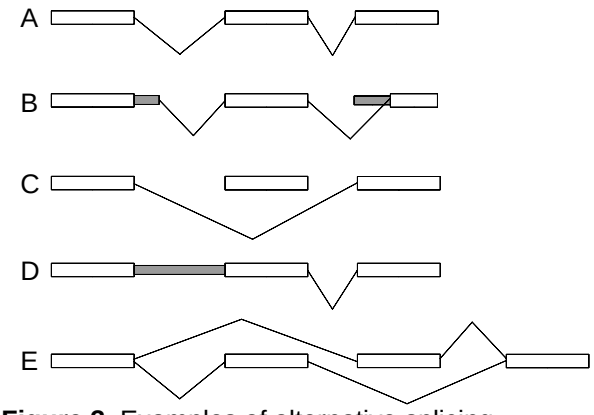

Figure 3. Examples of alternative splicing patterns. White boxes, exons; grey boxes, alternative exonic sequences; black lines, splicing patterns. A. Constitutive splicing. B. Alternative 5' and 3' splice sites. C. Exon skipping. D. Intron retention. E. Mutually exclusive alternative exons.

1.5.2 Regulation of alternative splicing

Use of alternative splice sites is influenced by the competition between SR

proteins and hnRNP proteins, the former having a predominantly enhancing role, whereas the latter are usually inhibitory. Members of the SR protein family share common structural features, including the arginine/serine-rich (RS) domain and at least one RNA recognition motif (Zahler *et al.* 1992; reviewed by Graveley 2000). Other proteins with an RS domain, many of which function in splicing, are nominated SR-like proteins. The SR protein SF2/ASF (splicing factor 2/alternative splicing factor) and hnRNP A1 are examples of antagonistic splicing factors (Mayeda and Krainer 1992). SR protein binding sites are found in most exons and are often required also for constitutive splicing. In addition to factors binding to splicing enhancers and repressors, the polypyrimidine tract binding protein, also known as hnRNP I, is implicated in the repression of alternative exons (García-Blanco *et al.* 1989; reviewed by Valcárcel and Gebauer 1997). In addition to ubiquitous splicing regulators, many tissue-specific alternative factors have been found, e. g. the neuronal splicing factor Nova (Jensen *et al.* 2000) and Fox1 that functions in both brain and muscle tissue (Jin *et al.* 2003). The additive effects of different regulatory elements and the factors binding to them determine the splicing pattern of the mRNA (see e. g. Han *et al.* 2005). Through this competitive action, weak splice sites can be used or ignored.

Alternative splicing events have been studied on the whole transcriptome level using microarrays with exon, exon junction or tiling probes (Johnson *et al.* 2003, Pan *et al.* 2004, Castle *et al.* 2008; reviewed by Hallegger *et al.* 2010). Array-based methods are limited to relatively abundant mRNA species and tend to miss rare events. Recently, high-throughput sequencing has become an attractive method of detecting alternative splicing (Wang *et al.* 2008, Richard *et al.* 2010). Attempts have been made to assemble a general splicing code that could be used to predict alternative splicing events based on sequence (Barash *et al.* 2010). The task is complicated by the redundancy and complexity of the elements and factors enhancing and inhibiting splice site recognition.

Alternative splicing can also be regulated by chromatin modifications, and inversely, gene structure can influence chromatin packaging. Nucleosome density has been found to differ between introns and exons (Spies *et al.* 2009, Chen *et al.* 2010). Histone modifications can affect alternative splicing by influencing the speed of the RNA polymerase or the recruitment of splicing factors to the nascent pre-mRNA (Schor *et al.* 2009, Luco *et al.* 2010).

1.5.3 Alternative splicing by the U12-type spliceosome

Alternative splicing patterns rarely involve U12-type introns, most probably because of the longer and more rigid splicing consensus sequences required by the U12-type spliceosome. Also, the co-operative intron recognition by the U11/U12 di-snRNP (see **1.9.3**) reduces flexibility in splice site recognition. Use of alternative 3' splice sites in U12-type introns has been observed, but the question whether they represent true alternative splicing or errors has not been resolved (Levine and Durbin 2001, Zhu and Brendel 2003). In some cases, an intron can be spliced by either the U2- or the U12-type spliceosome at slightly differing splice sites. The Drosophila gene *prospero* contains a double intron, 'twintron', consisting of a U2-type intron nested inside a U12-type one (Hall and Padgett 1994). The splicing of the twintron is regulated by a purine-rich element, which binds Drosophila homologues of hnRNP A1, Hrp38 and Hrp36 proteins, influencing the function of the U2-type and, to a lesser degree, the U12-type spliceosome (Scamborova *et al.* 2004, Borah *et al.* 2009). In some cases a U12-

type intron is retained in a considerable proportion of the mRNA pool (**I**). Whether intron retention is the consequence of splicing regulation or just a side-effect of the low efficiency of the U12-dependent spliceosome is unclear. Generally, U12-type introns have an effect on the expression levels rather than on the content of the mRNAs they reside in.

1.5.4 Other forms of regulated splicing

Trans-splicing forms yet another way of creating variation to the transcriptome. This means of combining exons in different mRNA molecules was first described in trypanosomes, where a capped leader sequence is joined to the 5' end of transcripts, and in plant organelles (reviewed by Bonen 1993). Trans-splicing events have since been reported to occur infrequently also in some nuclear mRNAs in animals (Dorn *et al.* 2001).

Regulation of the level of splicing efficiency is rare, but in yeast, the splicing of several pre-mRNAs encoding for ribosomal proteins is inhibited as a response to amino acid starvation (Pleiss *et al.* 2007).

1.6 Splicing and disease

1.6.1 Aberrant splicing has a role in many diseases

Aberrant splicing can play an important role in disease (for review, see Faustino and Cooper 2003, Wang and Cooper 2007). Mutations that cause splicing defects can be divided in three categories. First, *cis*-acting mutations may arise in splicing consensus sequences or splicing enhancers or inhibitors. In a mutation analysis performed on the cystic fibrosis transmembrane conductance regulator gene, about one quarter of synonymous mutations in exons caused exon skipping (Pagani *et al.* 2005). It has been estimated that more than half of known disease-causing mutations disrupt splicing (López-Bigas *et al.* 2005). Second, *trans*-acting mutations may affect the activity of factors required for constitutive or alternative splicing. Third, the function of common spliceosome components can be impaired (see **1.6.2**). The first two types of mutations cause erroneous splicing that may result in the inappropriate exclusion of part of the coding sequence, the inclusion of ectopic residues in the polypeptide produced, or the introduction of a premature stop codon. A change in splicing isoform ratios may also have serious effects. Abnormal ratios of alternative isoforms of the tau protein cause its aggregation and play a role in a variety of human neurodegenerative disorders such as Alzheimer's disease (reviewed by Liu and Gong 2008). Therapeutic approaches have been made to correct splicing mutations using e. g. modified oligonucleotides to mask cryptic splice sites (Garcia-Blanco *et al.* 2004). Ratios of splicing variants can also be used as tools for diagnostics, e. g. to monitor drug efficacy in Alzheimer's disease (Darreh-Shori *et al.* 2004).

Splicing abnormalities are common in cancer. The SR protein ASF/SF2 is upregulated in many cancers and promotes the production of oncogenic isoforms of the ribosomal protein S6 kinase-β1 and the Ron protooncogene (Ghigna *et al.* 2005, Karni *et al.* 2007). Alternative splicing has implications for the severity and prognosis of cancer, and cancer-specific transcripts can serve as biomarkers and diagnostic tools (Dutertre *et al.* 2010).

1.6.2 Mutations in spliceosome components may have tissue-specific phenotypes

Diseases associated with mutations in core spliceosome components do not always result in pleiotrophic effects but may elicit tissue-specific phenotypes. Mutations in the

spliceosomal PRP31, PRP3 and PRP8 proteins, all components of the U4/U6.U5 (and U4atac/U6atac.U5) snRNP complex, cause retinitis pigmentosa, a disease characterized by retinal degeneration and ultimately blindness (McKie *et al.* 2001, Vithana *et al.* 2001, Chakarova *et al.* 2002, Makarova *et al.* 2002). The underlying reason for the tissue-specificity of symptoms caused by a mutation in a common splicing factor is unknown. Possible explanations include a requirement of high levels of one or a few mRNAs in the affected tissue, or the sensitivity of tissue-specific genes to any change in the stoichiometry of splicing factors (Faustino and Cooper 2003).

Spinal muscular atrophy (SMA), one the most common genetic causes of childhood lethality, is caused by mutations in the gene encoding for the survival of motor neurons (SMN) protein (Lefebvre *et al.* 1995; for review, see Burghes and Beattie 2009). SMA is caused by the loss of spinal cord motor neurons and results in progressive muscle weakness and paralysis. SMN has an essential function in the biogenesis of spliceosomal snRNPs in facilitating the binding of the Sm protein complex to snRNAs (Fischer *et al.* 1997; see **1.7.1**, **1.8.1**). SMN deficiency causes a reduction in the levels of spliceosomal snRNPs and, interestingly, has a stronger effect on minor snRNPs, particularly U11 (Bühler *et al.* 1999, Gabanella *et al.* 2007). SMA patients have splicing defects, but the sequence of pathological events associated with SMA has not yet been deciphered (Zhang *et al.* 2008). In the absence of SMN, Sm proteins can bind inappropriate RNAs, which may contribute to the disease (Pellizzoni *et al.* 2002).

1.7 Spliceosomal composition

1.7.1 Components of the U2-type spliceosome

The core components of the U2-type spliceosome are for the most part conserved from yeast to human. The major splicing pathway involves five snRNAs, U1, U2, U4, U5 and U6, assembled into snRNPs that are named after the snRNAs (Hodnett and Busch 1968, Reddy *et al.* 1974, Lerner and Steitz 1979, Krämer *et al.* 1984, Krainer and Maniatis 1985, Black *et al.* 1985, Chabot *et al.* 1985, Black and Steitz 1986). U1, U2, U4 and U5 contain a conserved uridine-rich sequence flanked by stem loops, the Sm site, that is bound by the seven Sm proteins B/B', D1, D2, D3, E, F and G (Branlant *et al.* 1982, Liautard *et al.* 1982, Bringmann and Lührmann 1986, Lehmeier *et al.* 1990). Besides the common Sm proteins, each snRNP contains its own specific proteins. U6 snRNP does not contain Sm proteins but is instead assembled with the Sm-like proteins Lsm2-8 (Séraphin 1995, Salgado-Garrido *et al.* 1999, Achsel *et al.* 1999). The spliceosomal core proteins are summarized in Table 1.

A number of DExD/H box proteins, many of which possess ATPase and RNA helicase activity, such as the U5-200K and U5-100K proteins, provide energy for the conformational changes required for spliceosome formation (see **1.9.2**; reviewed by Staley and Guthrie 1998). The yeast orthologue of U5-100K, Prp28p, has been implicated in disrupting the interaction between U1 snRNA and the 5′ splice site (Teigelkamp *et al.* 1997, Staley and Guthrie 1999). Brr2p/U5-200K may be involved in disrupting the U4/U6 base pairing (Laggerbauer *et al.* 1998, Raghunathan and Guthrie 1998b). Other snRNP-specific proteins include a putative GTPase, the ribosomal elongation factor 2-like U5-116K protein (Fabrizio *et al.* 1997). The single U6-specific protein p110/SART3 (Prp24 in yeast) is a recycling factor involved in

restoring the base pairing between U4 and U6 (Raghunathan and Guthrie 1998a, Bell *et al.* 2002).

In addition to Sm and snRNP-specific proteins, numerous non-snRNP proteins are involved in splicing. These include RNA binding proteins that facilitate the formation of RNA-RNA interactions, such as U2AF-65 that anneals U2 with the branch point sequence (Valcárcel *et al.* 1996). Up to about 300 different proteins have been identified by purification and mass spectrometry sequencing of spliceosomal complexes at different stages of splicing, making the spliceosome the largest intracellular machine (Hartmuth *et al.* 2002, Jurica *et al.* 2002, Rappsilber *et al.* 2002, Zhou *et al.* 2002, Deckert *et al.* 2006; reviewed by Jurica and Moore 2003). However, the identification of true splicing factors from other RNA-binding proteins is not straightforward. Due to the importance of splicing enhancers to the processing of mRNA in higher eukaryotes, the splicing of many pre-mRNAs is likely to require auxiliary proteins binding to exonic and intronic splicing enhancers in addition to the spliceosome core factors. Given the degeneracy of the U2-type splicing consensus sequences, additional factors may be more numerous in the case of U2-type than U12-type introns. Some protein factors associated with the spliceosome are involved in communication with other steps in mRNA maturation, such as the transcription complex, capping and polyadenylation machinery and the exon junction complex (EJC; see **1.10**).

1.7.2 Components of the U12-type spliceosome

The overall composition of the U12-dependent spliceosome is similar to the major spliceosome. The snRNAs U1, U2, U4 and U6 are replaced by their functional analogues U11, U12, U4atac and U6atac, respectively, whereas U5 is shared between the two systems (Montzka and Steitz 1988, Hall and Padgett 1994, Tarn *et al.* 1995, Hall and Padgett 1996, Tarn and Steitz 1996a, Tarn and Steitz 1996b, note a corrected U4atac/U6atac secondary structure in Padgett and Shukla 2002). The snRNAs specific to the U12-type spliceosome are structurally similar to and share some sequence features with their major counterparts (Figure 4). The corresponding snRNPs U11, U12 and U4atac contain the Sm protein complex, and U6atac is assembled with the Lsm proteins. Similar to U4 and U6, U4atac and U6atac interact by base pairing and associate with U5 to form the minor tri-snRNP U4atac/U6atac.U5. Unlike U1 and U2, which exist mainly as monoparticles, U11 and U12 form a di-snRNP through protein-protein interactions (Wassarman and Steitz 1992b). The U11/U12 di-snRNP lacks U1-specific proteins but contains most proteins present in U2 snRNP, such as the SF3b complex (see Table 1). In addition, U11/U12 snRNP contains seven unique proteins not found in the U2-dependent spliceosome, designated 65K, 59K, 48K, 35K, 31K, 25K, and 20K, all of which are essential for cell viability (Will *et al.* 1999, Will *et al.* 2004). 59K, 48K, 35K and 25K are detected also in U11 monoparticles. 65K binds U12 snRNA and U11-59K protein, thus bridging U11 and U12 snRNPs (Benecke *et al.* 2005). Based on homology to U1-70K, 35K is believed to function in 5' splice site recognition (Will *et al.* 1999). 48K protein binds the 5' splice site transiently at the intron recognition stage, interacts with U11-associated 59K and has a role in U11/U12 di-snRNP stability (Turunen *et al.* 2008, Tidow *et al.* 2009). Furthermore, 59K protein (also known as ES18), has been implicated in apoptosis according to Park *et al.* (1999), but in their study it was assumed to be a transcription factor. In contrast to the U11/U12 di-snRNP, the protein components of the U6atac/U4atac.U5 tri-snRNP

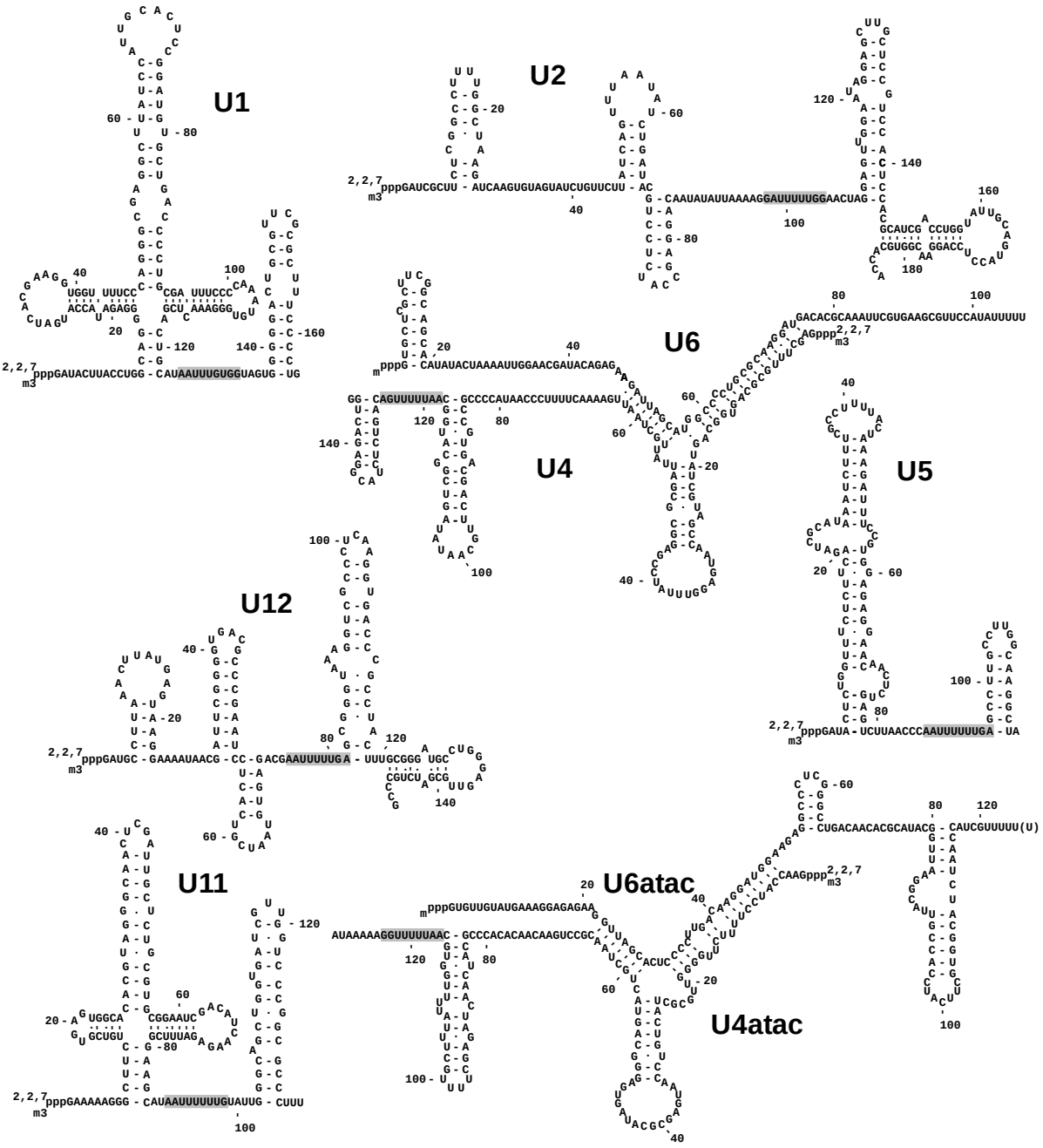

Figure 4. Predicted schematic secondary structures of human spliceosomal snRNAs. The Sm complex binding site is marked with grey shading. The structures of U1, U2, U4, U5 and U6 after Yu *et al*. (1999), U11 and U12 after Tarn and Steitz (1997), U4atac and U6atac after Padgett and Shukla (2002).

apparently do not differ from those in the major tri-snRNP (Luo *et al.* 1999, Nottrott *et al.* 2002, Schneider *et al.* 2002).

U12-type spliceosome components are about 100 times less abundant than those of the U2-type spliceosome (Montzka and Steitz 1988, Tarn and Steitz 1996a). Due to the large number of U snRNA pseudogenes and the difficulty of sequencing tandemly repeated short genes, the exact number of functional snRNA genes is not known. The copy number estimates

of U2-type spliceosome snRNA genes range from a few to several tens of copies, whereas minor spliceosome snRNAs are encoded by single-copy genes in most organisms (Lander *et al.* 2001, Marz *et al.* 2008). The expression of different major snRNA variants changes during development, but the significance of the variants is poorly understood (Forbes *et al.* 1984, Lund *et al.* 1985, Hanley and Schuler 1991, Domitrovich and Kunkel 2003).

Spliceosomal composition is highly conserved among eukaryotes and especially in multicellular organisms, but some differences are eminent. Drosophila has orthologues to most human spliceosomal components, and the dynamics of entry and release of factors during spliceosome assembly and splicing reactions appear conserved between fly and vertebrates (Mount and Salz 2000, Herold *et al.* 2009). The fly U11 snRNA is markedly diverged on sequence level, with only 28% similarity to human U11, but similarities in secondary structure and complementarity to the 5' splice site suggest conservation of function (Schneider *et al.* 2004). Also the U11/U12 proteins 65K and 20K are highly diverged in Drosophila, and homologues to U11-specific proteins have not been found. However, homologues to U11-35K protein have been found in other insects (Will *et al.* 2004).

Table 1. Human spliceosomal snRNP proteins.

snRNP	Shared proteins	Spliceosome-specific proteins
U1	Sm proteins[1]	U1-70K, U1-A, U1-C
U2	Sm proteins[1], SF3b[2]	SF3a[2], U2-A', U2-B"
U5	Sm proteins[1], U5-220, 200, 116, 102, 100, 52, 40, 15K	-
U4/U6, U4atac/U6atac	Sm proteins[1], Lsm2-8, U4/U6-90, 61, 60, 20, 15.5K	-
U4/U6.U5, U4atac/U6atac.U5	Sm proteins[1], Lsm2-8, U5- and U4/U6-specific proteins, 27K, 65K, 110K	-
U11/U12	Sm proteins[1], SF3b[2]	[3]65K, 59K, 48K, 35K, 31K, 25K, 20K

After Burge *et al.* (1999), Will *et al.* (1999), Schneider *et al.* (2002), Will *et al.* (2004)
[1]Sm proteins B/B', D1, D2, D3, E, F and G
[2]Multi-subunit complexes

1.7.3 Evolutionary origin of the two spliceosomes

The origins of the two parallel splicing systems remain elusive, and several hypotheses have been presented. The overall similarities in the composition of the spliceosomes, the secondary structures and functions of the snRNAs and the large number of shared proteins suggest a common origin. The endosymbiont theory postulates that a unicellular ancestor of eukaryotes, already in possession of introns and a spliceosome, split into two lineages which continued to diverge, accumulating differences in spliceosomal components and intron consensus sequences (Burge *et al.* 1998). Later, these two lineages fused, possibly by endosymbiosis, to give rise to the ancestor of present-day eukaryotes. Another hypothesis suggests that the two spliceosomes have developed through parallel evolution after

duplication (Tarn and Steitz 1997). In the latter case, U12-type introns should be randomly distributed in the genome. Instead, they tend to reside in genes with functions related to information processing, with an overrepresentation of genes that have two U12-type introns. The endosymbiont theory seems to provide a more plausible explanation to the similarities and differences between the two spliceosomal systems. During the separation of the spliceosomes, the snRNAs may have evolved faster than proteins, because their functionality is more dependent on secondary structure than actual sequence. After fusion, many less diverged components would then have been replaced by their counterparts in the other spliceosome. All the proteins found so far that are specific to the U12-type spliceosome are constituents of the U11/U12 di-snRNP, implying that the major differences between the two systems occur at the intron recognition step (see **1.9.3**).

1.8 Biogenesis of spliceosomal snRNPs

1.8.1 Biogenesis of Sm-class snRNPs

In multicellular organisms, spliceosomal snRNPs transiently localize in the cytoplasm during their biogenesis (reviewed by Will and Lührmann 2001, Patel and Bellini 2008). Factors required for the nuclear export of snRNAs or the re-import of snRNPs have not been found in *S. cerevisiae*, suggesting that snRNP biogenesis is entirely nuclear in yeast (Huber *et al.* 1998, Ohno *et al.* 2000). The snRNPs of the U12-type spliceosome are believed to follow the same maturation path as their counterparts in the major spliceosome.

After transcription by the RNA polymerase II, the Sm-class snRNAs U1, U2, U4 and U5 acquire a monomethylated cap and are exported from the nucleus. Export requires the monomethyl cap structure and is mediated by the export adapter PHAX specific to U snRNAs (Hamm and Mattaj 1990, Ohno *et al.* 2000). In the cytoplasm, the seven Sm proteins are assembled onto the Sm site. In metazoa, the assembly of the Sm protein ring is facilitated by the SMN complex consisting of the SMN protein, seven Gemin proteins (Gemins 2-8) and several other factors (Pellizzoni *et al.* 2002). After the Sm core assembly, the cap is modified into a 2,2,7-tri-methyl-guanosine (m3G) structure (Mattaj 1986). The tri-methyl cap and the Sm complex are nuclear localization signals that cause snRNPs to be subsequently re-imported into the nucleus (Fischer and Lührmann 1990, Fischer *et al.* 1993). Nuclear import is mediated by the import adapter snurportin1 (Huber *et al.* 1998). SMN protein interacts with snurportin1 and is imported into the nucleus in complex with snRNPs (Narayanan *et al.* 2002, Narayanan *et al.* 2004). After nuclear re-import, snRNPs acquire their specific proteins and are internally modified by pseudouridylation and 2'-O-methylation by a mechanism guided by small RNAs (Yu *et al.* 2001). Some controversy remains over the location of these modifications. Newly assembled snRNPs accumulate in subnuclear organelles termed Cajal bodies (coiled bodies), where small Cajal body-specific RNAs that guide snRNA modifications are also enriched (Sleeman and Lamond 1999, Jády *et al.* 2003). Also the nucleolus has been suggested as the site of further processing (Yu *et al.* 2001). Cajal bodies may also be the sites of assembly and recycling of the U4/U6.U5 tri-snRNP (Stanek *et al.* 2003, Schaffert *et al.* 2004). Mature snRNPs are enriched in interchromatin granule clusters aka speckles, which are believed to be storage sites for splicing factors (Matera and Ward 1993; reviewed by Lamond and Spector 2003).

The final step in the maturation of spliceosomal snRNPs is the formation of preassembled snRNP complexes. The major snRNPs U4 and U6 interact by base pairing to form the U4/U6 complex and associate with U5 by protein interactions forming the tri-snRNP U4/U6.U5 (Bringmann *et al.* 1984, Hashimoto and Steitz 1984, Cheng and Abelson 1987, Konarska and Sharp 1987). The minor spliceosome tri-snRNP U4atac/U6atac.U5 forms similarly (Tarn and Steitz 1996a, Tarn and Steitz 1996b). In addition, U11 and U12 associate to form a di-snRNP (Montzka and Steitz 1988, Wassarman and Steitz 1992b).

1.8.2 Biogenesis of U6 and U6atac

The biogenesis of U6 and presumably U6atac snRNPs differs from that of the other spliceosomal snRNPs. U6 is transcribed by the RNA polymerase III (Kunkel *et al.* 1986) and is believed not to be exported from the nucleus. The polyuridine tail of nascent U6 is transiently bound and stabilized by the La protein (Rinke and Steitz 1985, Pannone *et al.* 1998). The triphosphate cap is modified into a gamma-monomethyl phosphate cap (Shimba and Reddy 1994). As the other spliceosomal snRNAs, U6 is also modified by pseudouridylation and 2'-O-methylation by small nucleolar RNPs, probably in the nucleolus (Tycowski *et al.* 1998). U6 and U6atac lack an Sm binding site and instead, they are assembled with seven Sm-like proteins, the Lsm2-8 complex, which replaces the La protein and serves also as a nuclear retention/localization signal (Achsel *et al.* 1999, Schneider *et al.* 2002, Spiller *et al.* 2007).

1.9 Spliceosomal catalysis and assembly

1.9.1 Spliceosomal catalysis

Splicing is achieved by two phosphoryl transfer reactions (Padgett *et al.* 1984, Ruskin *et al.* 1984, Maschhoff and Padgett 1993, Moore and Sharp 1993). First, the 2′ hydroxyl of the branch site nucleotide attacks the phosphate at the 5′ splice site, producing a free 5′ exon and an intermediate composed of the 3' exon and a lariat intron. In the second reaction, the 3′ hydroxyl of the 5′ exon attacks the phosphate at the 3′ splice site, yielding a ligated mRNA and a lariat intron. These reactions conserve the energy of the phosphate bonds and do not require external energy input.

It is likely that splicing reactions are catalyzed by RNA, making the spliceosome a ribozyme (Kruger *et al.* 1982, Cech 1986; reviewed by Valadkhan 2007). The theory is corroborated by several lines of evidence. The structures formed by the RNA molecules in the catalytic core of the spliceosome resemble the conformation of self-splicing group II introns. U6 snRNA is situated in the catalytic core and coordinates metal ions (Yean *et al.* 2000). U2 and U6 snRNAs are capable of catalyzing a splicing-like reaction free of proteins *in vitro* (Valadkhan and Manley 2001, Valadkhan *et al.* 2009). The requirement for a vast number of proteins in the spliceosome despite RNA-based catalysis is thought to serve to maintain unidirectionality of the reactions, increase fidelity and speed and provide flexibility needed in alternative splicing.

The similarity of the splicing mechanisms and catalytic core structures of self-splicing group II introns with those of the spliceosome has promoted the hypothesis that spliceosomal introns have evolved from self-splicing introns (Sharp 1985, Toor *et al.* 2008). According to this theory, the spliceosome has originated through the dispersion of the self-splicing intron

ribozyme core into several small RNAs (Jarrell *et al.* 1988, Sharp 1991). Further support for the model is provided by the ability of the catalytic domain of a self-splicing group II intron to functionally replace the metal binding stem loop in U6atac (Shukla and Padgett 2002). With its most likely RNA-based catalytic mechanism, the spliceosome is an RNP machine, reminiscent of an ancient 'RNA world' before proteins took over most enzymatic functions in cells.

1.9.2 U2-type spliceosome assembly

The assembly pathway of the major spliceosome is conserved from yeast to mammals. In the stepwise assembly model, the assembly of the major spliceosome begins with the recognition of the splicing consensus sequences by U1 snRNP and protein factors, with U1 snRNA base pairing to the 5’ splice site, SF1 (splicing factor 1) binding to the branch point sequence and the U2AF subunits 65K and 35K to the polypyrimidine tract and the 3' splice site region, respectively, forming the E (early) complex (Mount *et al.* 1983, Black *et al.* 1985, Zamore and Green 1989, Abovich and Rosbash 1997; for review, see Staley and Guthrie 1998, Smith *et al.* 2008). On yeast introns, which lack a polypyrimidine tract, the analogous commitment complex is assembled through the recognition of the 5' splice site by U1 snRNP and the branch region by the SF1 orthologue BBP and the U2AF orthologue Mud2p (Berglund *et al.* 1997, Abovich and Rosbash 1997, Merendino *et al.* 1999, Zorio and Blumenthal 1999). SF1 and U2AF-35, together with SR proteins, are involved in bridging the intron ends by protein interactions. Complex A, or prespliceosome, is formed by the replacement of SF1 by U2 snRNP, which binds the branch point with protein-protein and base pairing interactions, bulging the branch site adenosine out of the duplex (Parker *et al.* 1987, Query *et al.* 1994). The stable binding of U2 requires ATP hydrolysis. The next stage, complex B, is formed upon the entry of the U4/U6.U5 tri-snRNP. Rearrangements of both RNA-RNA and protein interactions result in the expelling of U4 snRNP, the replacement of U1 by U6 at the 5' splice site and the formation of the catalytic core by U2 and U6 in the activated complex B* (Konarska and Sharp 1987, Sawa and Abelson 1992, Wassarman and Steitz 1992a). This reorganization is propelled by ATP-hydrolyzing proteins with RNA helicase activity (reviewed by Staley and Guthrie 1998). U6 forms two base paired helices with U2, bringing the 5' splice site and the branch point into close proximity for the first step of splicing (Hausner *et al.* 1990, Madhani and Guthrie 1992). U5 contacts the 5' splice site on the exon side and the 3' splice site and helps align the exons following the first step of splicing (Newman and Norman 1992, Sontheimer and Steitz 1993). The protein composition of the activated spliceosome is also remodelled at this stage (Makarov *et al.* 2002). The next stage, complex C, contains splicing intermediates subsequent to the first step of splicing reactions. The sequential recognition of each splicing signal by several components contributes to splicing fidelity.

In vitro, spliceosome assembly occurs in the above described stepwise manner. Some reports suggest that the spliceosome may exist as a preassembled complex. Complexes consisting of all five spliceosomal snRNPs have been isolated in yeast and HeLa cell nuclear extracts (Stevens *et al.* 2002, Malca *et al.* 2003). However, a preformed penta-snRNP is not a prerequisite for splicing *in vitro* (Behzadnia *et al.* 2006).

1.9.3 U12-type spliceosome assembly

The assembly pathway of the U12-dependent spliceosome resembles the major spliceosome, and interactions are functionally conserved (Kolossova and Padgett 1997, Yu and Steitz 1997, Incorvaia and Padgett 1998; see Figure 5). However, the initial intron recognition steps differ from those involving major introns. Unlike U1 and U2 in the major spliceosome, U11 and U12 snRNPs form a di-snRNP and recognize the 5' splice site and branch point cooperatively (Frilander and Steitz 1999). The base pairing interactions of U1 with the 5' splice site in the major spliceosome span nucleotides -1 to +6 relative to the 5' splice site, whereas U11 base pairs only to positions +4 to +8. The three first nucleotides of the 5' splicing consensus are recognized by one of the proteins specific to the U12-type spliceosome, the U11-48K protein (Turunen *et al.* 2008). Similarly to the assembly of the major spliceosome, the activated B complex is formed through the entry of the tri-snRNP U4atac/U6atac.U5, dissociation of U11 and U4atac and base pairing of U6atac with U12 and the 5’ splice site (Tarn and Steitz 1996a, Yu and Steitz 1997, Frilander and Steitz 2001). Important molecular structures and interactions are also conserved, as can be deduced from the preservation of splicing activity when the conserved intramolecular stem loop in U6atac is replaced with the corresponding structure in U6 (Shukla and Padgett 2001). Furthermore, U4atac can be replaced by U4 if the necessary base pairing interactions with U6atac are introduced (Shukla and Padgett 2004).

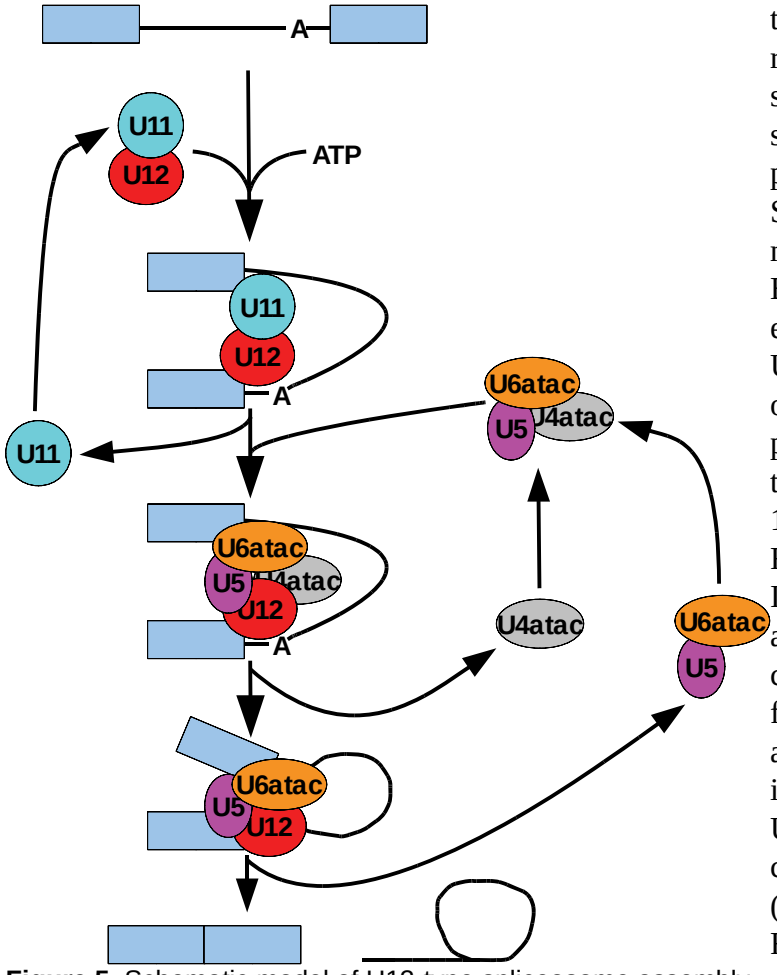

Figure 5. Schematic model of U12-type spliceosome assembly. Blue boxes, exons; black line, intron. 'A' designates the branch point nucleotide.

1.9.4 Spliceosome recycling

After the completion of splicing reactions, the spliceosome must be disassembled and the components recycled for the next round of splicing. Several proteins have been implicated in recycling, many of which also have a role in the initial assembly of spliceosomal snRNPs. These include p110/SART3/hPrp24 and hPrp31 (U4/U6-61K), which are essential to the recycling of both U4/U6 and U4atac/U6atac complexes. They are enriched in Cajal bodies,

which are suggested as the sites of tri-snRNP reassembly (Schneider *et al.* 2002, Stanek *et al.* 2003, Damianov *et al.* 2004, Schaffert *et al.* 2004, Stanek *et al.* 2008).

1.10 Communication with other processing systems

1.10.1 Cotranscriptional splicing

Although mRNA splicing can occur uncoupled from transcription, both U2- and U12-type splicing is normally carried out or at least initiated during transcription (Beyer and Osheim 1988, Singh and Padgett 2009). Cotranscriptional processing is more efficient (Bird *et al.* 2004, Das *et al.* 2006). The exact mechanisms of spliceosome recruitment are not well known. In yeast, U1 is recruited to the transcript cotranscriptionally (Kotovic *et al.* 2003). In metazoa, phosphorylation of the carboxy-terminal domain of the RNA polymerase II is essential for cotranscriptional splicing (Bird *et al.* 2004). Splicing is conducted in concert with other post-transcriptional processing steps in a network of mutually enhancing interactions (reviewed by Reed 2003, Moore and Proudfoot 2009). Splicing activity is enhanced by the processing of adjacent introns, and pre-mRNA processing stimulates transcription (Fong and Zhou 2001). Polyadenylation promotes splicing of the terminal exon, and splicing in turn increases the rate of polyadenylation (Niwa *et al.* 1990, Niwa and Berget 1991). Nascent transcripts are retained at the transcription site until the removal of introns has been completed (Custódio *et al.* 1999). Splicing also promotes the export of the mRNA out of the nucleus (Chang and Sharp 1989, Valencia *et al.* 2008). Transgenes from which introns have been removed often fail to be expressed (Hamer and Leder 1979). In some cases, introns can be retained until later or even spliced in the cytoplasm of some cell types, e.g. in platelets lacking a nucleus, or neuronal dendrites (Denis *et al.* 2005, Glanzer *et al.* 2005).

1.10.2 mRNA quality control by splicing

Splicing also has an important function in the control of mRNA quality. Subsequent to splicing, an exon junction complex is deposited at the site of the removed intron (Kataoka *et al.* 2000, Le Hir *et al.* 2000). In nonsense-mediated decay, transcripts from genes that have acquired a premature termination codon as a result of mutation are recognized based on the presence of an EJC downstream of a stop codon and targeted to degradation (reviewed by Conti and Izaurralde 2005). EJC formation is conserved between U2- and U12-type spliceosomes (Hirose *et al.* 2004). Failure to splice an intron would then lead to the degradation of the mRNA, which is believed to happen to pre-mRNAs with inefficiently spliced U12-type introns.

2 Aims of the study

This work focuses on the specific characteristics of the U12-type spliceosome, in which its function differs from the U2-type system. In particular, the low efficiency of U12-type splicing is believed to have important consequences to gene expression by limiting the production of mature mRNAs from genes containing U12-type introns.

The inefficiency of U12-type splicing has been attributed to the low abundance of the components of the U12-type spliceosome in cells, but this hypothesis has not been proven. The aim of the first part of this work was to study the effect of the abundance of the spliceosomal snRNA components on splicing. U4atac snRNA was chosen as the target based on previous reports, which led to the hypothesis that U4atac levels might be limiting the activity of the U12-type spliceosome.

The inefficiency of U12-type splicing has also promoted the idea that the U12-type spliceosome controls gene expression, limiting the mRNA levels of some U12-type intron-containing genes. While the identities of the primary target genes are relatively well known, little has previously been known about the downstream genes and pathways possibly affected by this regulation. Splicing mutant *Drosophila* was studied with the aim of screening the effects of impaired U12-type splicing to the transcriptome of a whole organism.

The subcellular localization of the U12-type spliceosome was studied with the aim of settling a recent dispute on whether U12-type splicing would be localized in the cytoplasm. The answer to this fundamental question was crucial to the field.

3 Materials and methods

Methods used in this study are listed in Table 2. For a detailed description of methods, see the original publications.

Table 2. Methods used in this work.

Method	Article
Cell lines and culture	I, II
Fly strains	III
Expression plasmids	I
RNA isolation	I, II, III
Northern blotting	I, II, III
Native gel analysis	I
Psoralen cross-linking	I
RT-PCR	I, III
Quantitative RT-PCR	I, III
Nonradioactive *in situ* hybridization	II
Fluorescent *in situ* hybridization	II
Cell fractionation	II
Western blotting	II
Microarray analysis	III

4 Results and discussion

4.1 The abundance of U12-dependent spliceosome components does not limit endogenous splicing efficiency (I)

4.1.1 U4atac is the limiting component in the formation of U4atac/U6atac complex

The efficiency of U12-type splicing has been observed to be lower than normal splicing *in vitro* and *in vivo* on transfected constructs (Patel *et al.* 2002). The inefficiency of U12-type splicing has been attributed either to the low abundance of U12-type spliceosome components or to a slow rate of catalysis. In this study, the former hypothesis was addressed. The expression levels of spliceosomal snRNAs in mouse NIH-3T3 and L-929 and human HeLa cell lines were studied by Northern blotting. Large variation in the expression levels of U4atac snRNA between different cell lines suggested that its levels could be regulated, while the levels of the other snRNAs did not exhibit much variation. Therefore it was of interest to find out whether changes in U4atac levels would have an effect on the formation of spliceosomal higher-order complexes and eventually on splicing efficiency. Since U4atac and U6atac apparently are the only components of the U4atac/U6atac.U5 tri-snRNP that are specific to the U12-type spliceosome (Schneider *et al.* 2002), and the components shared with the major spliceosome are much more abundant, we hypothesized that U4atac may also limit the abundance of the minor tri-snRNP. If so, U4atac levels could directly influence splicing efficiency. At least *in vitro*, the intron recognition steps prior to tri-snRNP entry are faster than the following events leading to spliceosome activation (Tarn and Steitz 1996a, Tarn and Steitz 1996b, Frilander and Steitz 1999, Frilander and Steitz 2001), which suggests that the bottleneck of U12-type splicing occurs at this stage. On the other hand, *in vitro* splicing may not accurately reflect the situation *in vivo*.

To test the hypothesis that U4atac levels regulate the activity of the U12-type spliceosome by limiting the amount of U4atac/U6atac complex, we studied a derivative of the mouse fibroblast cell line 3T3, dubbed 3T3-D1, in which the level of U4atac snRNA was reduced 5-fold compared to other cell lines tested, including L-929 cells that were used as control. Nuclear extract was prepared from 3T3-D1, L-929 and HeLa cell lines and RNA was isolated in conditions that preserved the base pairing between U6atac and U4atac. Samples were then analyzed on native gels followed by Northern blotting. Plenty of free U6atac not bound to U4atac was detected but the amount of free U4atac was negligible, also in cells containing normal U4atac levels. The U4atac/U6atac complex was reduced in 3T3-D1 cells, corresponding to the reduction in U4atac level. Therefore, U4atac is the limiting component in the formation of the U4atac/U6atac di-snRNA. The identities of the snRNAs in the complex were confirmed by RNA crosslinking analysis on nuclear extracts from the same cell lines.

4.1.2 Low U4atac levels inhibit splicing of a transfected construct

The effect of U4atac levels on splicing efficiency was studied using a splicing reporter construct, which consisted of a part of the mouse *smE* gene, containing one U12-type and one U2-type intron, fused with a fluorescent protein-coding sequence. The reporter was transiently transfected into 3T3-D1 cells, which have low U4atac levels, and control L-929 cells. Unspliced reporter pre-mRNA was detected by Northern blotting in 3T3-D1 cells, whereas

splicing in control cells was more efficient. The results were confirmed by RT-PCR, in which two products were detected. A large product corresponding to a reporter pre-mRNA with an unspliced U12-type intron and a correctly spliced normal intron was strongly amplified from the 3T3-D1 sample, but only weakly from the control sample. In contrast, a smaller product corresponding to the fully spliced mRNA dominated in control samples but was weaker in 3T3-D1. The identities of the PCR products were confirmed by cloning and sequencing.

To test whether the differences in splicing were due to U4atac abundance, we constructed a U4atac overexpression plasmid that contained the promoter of the mouse U12 snRNA gene fused to the coding part and downstream transcription terminator sequence of the U4atac gene. The splicing reporter was cotransfected together with the U4atac overexpression construct into 3T3-D1 and control cells and splicing was analyzed by Northern blotting and RT-PCR as above. Splicing of the reporter construct was markedly improved in 3T3-D1 cells. In control cells, a slight decrease in the signal from the unspliced pre-mRNA and an increase in the fully spliced mRNA were observed. The abundance of U4atac snRNA may thus limit the activity of the U12-type spliceosome in a situation where cells with low U4atac levels are burdened with high expression of U12-type introns.

4.1.3 Splicing of endogenous introns is not limited by U4atac levels

To find out whether U4atac levels have an effect on the splicing of endogenous genes, selection was applied on 3T3-D1 and L-929 cells transfected with the U4atac overexpression construct to obtain stably transfected cells. Normal function of the overexpressed U4atac snRNA was confirmed by the isolation and native gel analysis of snRNA complexes from transfected cells. Overexpressed U4atac behaved similarly to the endogenous snRNA and incorporated into the U4atac/U6atac complex, which was increased approximately 5-fold in 3T3-D1 cells. This was approximately equal to the amount of the complex in untransfected control cells.

The splicing efficiency of endogenous U12-type introns was measured by quantitative RT-PCR (qRT-PCR) in 3T3-D1 and control cells with and without stable U4atac overexpression. We chose a set of genes, *pex16*, *drap1*, *ipo4*, *psmc4*, *gars* and s*mE*, each with a U12-type intron representing either GT-AG or AT-AC subtypes and ranging from 85 to 1172 nt in length. Primers were designed for one U12-type and one U2-type intron in each gene so that 'intron' primers amplified a sequence across the exon-intron boundary, and one of the primers in each 'exon' primer pair spanned the exon-exon junction. The proportion of unspliced U12-type intron-containing pre-mRNA relative to spliced mRNA varied from less than 1% to 40% between genes, but a significant difference between 3T3-D1 and L-929 cells was observed only with *pex16*, which was spliced very inefficiently in 3T3-D1 cells. U4atac overexpression modestly decreased the U12-type intron signal from all genes tested in both 3T3-D1 and control cells, but had no effect on the levels of mature mRNA. We conclude that the abundance of the spliceosomal components has no significant effect on endogenous gene expression in normal situations.

4.1.4 U12-type introns are spliced less efficiently than normal introns

According to the qRT-PCR analysis of the six genes in 3T3-D1 and L-929 cells, the signal from the unspliced U12-type intron was consistently 2- to 3-fold higher than from the U2-type intron except for *psmc4*, which was spliced inefficiently at both U12-type and U2-

type junctions in L-929 cells. Thus, U12-type introns appear to be spliced less efficiently than normal introns. To find out whether expression levels correlate with splicing efficiency, we measured the levels of the mRNAs analyzed above by qRT-PCR. They exhibited 30-fold variation between genes, but no significant correlation was observed between expression level and splicing efficiency.

Also other reports (Patel *et al.* 2002) show that the U12-type spliceosome is not saturated by the expression of transfected U12-type introns at high levels. In this light, the presence of unspliced introns in normal cells is puzzling. The situation may vary between different cell types and tissues. It is also possible that we did not detect an effect of elevated U4atac level on splicing because transfected cells expressed U4atac at variable levels. The function of U4atac is to sequester U6atac until needed, and high U4atac levels might even be inhibitory to splicing. Yet another explanation to our results is that at elevated U4atac levels, some other component of the U12-type spliceosome becomes limiting. The phenomenon could also be explained by the ability of the U12-type spliceosome to respond to a higher demand, perhaps by detection of the amount of free components not engaged in splicing reactions in the nucleoplasm.

According to a recent report by Singh and Padgett (2009), U12-type introns are spliced almost as quickly after transcription as U2-type introns; normal introns were spliced approximately 5 minutes after transcription, and the lag time for U12-type introns was only twice as much. However, this experiment measured only the first appearance of the spliced product; it did not assess the average splicing rate. In the light of our data and other reports (see **4.1.1**) detecting inefficient U12-type splicing both *in vitro* and *in vivo* in various metazoan systems, it is likely that U12-type introns can serve to limit gene expression levels.

4.2 Deficient U12-type splicing causes perturbations in metabolic gene expression in *Drosophila* (III)

4.2.1 U6atac-deficient larvae die at 3rd larval stage

To find out more about the genes and pathways possibly affected by the regulation of U12-type splicing activity, we investigated the effects of a U12-dependent splicing defect in a whole organism. We chose the fruit fly as a model organism because it has only about 20 U12-type introns, which yet exhibit a similar preference to reside in information processing genes as in mammals (Schneider *et al.* 2004, Alioto 2007). Most *Drosophila* U12-type introns have an orthologue in mammals (Lin *et al.* 2010). Further benefits of the fly as a model organism are provided by the readily available genetic tools, particularly mutant strain collections.

We studied a fly strain deficient in U12-type splicing as described previously by Otake *et al.* (2002). This strain has a P-element-based transposon insertion in the gene coding for U6atac snRNA. In homozygotes, the mutation causes impaired U12-type splicing and the death of larvae at the 3rd larval instar. We began by studying the expression of wild-type and mutant U6atac in larvae by Northern blotting. As detected earlier by Otake *et al.* (2002), wild type U6atac snRNA was still present in 1st and 2nd instar homozygous larvae as maternal contribution, but by the 3rd instar, it disappeared and the expression of the mutant form was increased. It is not clear why the mutant U6atac was almost absent in heterozygotes. The mutated snRNA may be unable to compete with wild type U6atac on the binding of snRNP

proteins, thus becoming unstable. Heterozygous larvae had normal levels of wild type U6atac and a normal phenotype, so we used them as control samples in our experiments.

Analysis of the mutation phenotype was complicated by the location of the U6atac gene inside the intron of another gene, CG13394. The 10.7 kb transposon element inserted in the U6atac gene disturbed splicing of the host gene mRNA and prevented the inclusion of the alternative exon downstream. To assess the effect of the change in CG13394 expression on the U6atac mutation phenotype, we studied another mutant line, which contains a transposon element insertion in the first exon of CG13394. This line is viable and has a normal phenotype as a homozygote. We performed microarray analysis on the host gene mutant (as in **4.2.2**) and detected no significant expression changes in any genes including the mutated gene. RT-PCR analysis on the chimeric sequences of the mutant CG13394 gene showed that the transcription of CG13394 was not disturbed by the insertion mutation. Therefore, we conclude that the U6atac mutant phenotype is due to the disruption of U6atac function.

4.2.2 Genes with a U12-type intron show variable responses to defective splicing

Microarray analysis was conducted using a custom-designed cDNA array provided by Agilent Technologies. The array contained a 60 nt probe for most Drosophila exons annotated in the Ensembl database, excluding some very short or low-complexity exons. Samples were total RNA from 1st, 2nd and 3rd instar homozygous and heterozygous larvae.

Initial analysis indicated 2, 62 and 632 significantly changed genes in the 1st, 2nd and 3rd instar, 1, 23 and 416 of which were downregulated, respectively, at confidence level $p < 0.01$. Progressively increasing changes reflect the decline of wild-type U6atac levels in the homozygous mutant during larval development.

Genes containing a U12-type intron exhibited variable responses to the splicing defect. Of the 19 documented U12-type intron-containing genes studied, five showed significant upregulation of at least one probe in any larval stage, two had at least one downregulated probe and the remaining 12 exhibited no significant changes. The direction of the change did not correlate with expression level, position of the intron in the gene, or intron subtype (GT-AG or AT-AC). In general, all probes for a given U12-type intron-containing gene were changing in the same direction, even though for some probes, the change did not exceed the chosen threshold of significance. Possible changes in the splicing patterns of U12-type intron genes were studied on the level of individual exons. However, we did not detect any consistent effects on exon inclusion in the vicinity of U12-type introns.

The results for three U12-type intron-containing genes were confirmed by RT-PCR. FBgn0085478, FBgn0028703 (Nhe3) and FBgn0010551, were upregulated, nonsignificant and downregulated, respectively, according to array results. With RT-PCR using primers flanking a U12- or U2-type intron, accumulation of the unspliced U12-type intron signal was seen in each case in homozygous U6atac mutant, accompanied by a decrease in the level of spliced mRNA. The control amplicons across normal introns in FBgn0085478 and FBgn0028703 showed a modest upregulation, whereas FBgn0010551 control was slightly decreased, consistent with the array results.

In the case of U12-type intron-containing genes, upregulation of the array signal can be attributed to the accumulation of unspliced pre-mRNA and not to an increase in the level of fully spliced mRNA. On the other hand, according to RT-PCR, fully spliced mRNAs were

still present even in 3rd instar larvae. This is probably due to residual activity of the mutated U6atac in splicing, but could also result from trace amounts of maternal wild-type U6atac, or perhaps adjustment of the half-lives of the mRNAs to maintain their levels in spite of defective splicing. Thus, the levels of fully spliced mRNAs from upregulated U12-type intron genes are likely to be close to normal or modestly reduced in mutant larvae. In contrast, downregulation reflects true changes in mRNA levels. The variable responses of U12-type intron genes to the splicing defect may reflect differences in their susceptibility to expression level control by the U12-type splicing activity level. The downregulated mRNAs are most likely targeted to degradation as the result of inefficient splicing, whereas the other U12-type intron-containing genes may not be as sensitive to changes in splicing efficiency.

4.2.3 Downstream effects in U12-type splicing mutant may involve a mitochondrial defect

We then focused our attention on the downregulated genes to obtain clues on how the downstream effects could be explained. FBgn0010551 (l(2)03709), the most strongly downregulated U12-type intron gene, exhibited mild nonsignificant downregulation already at the first larval instar, suggesting that it was among the first genes that changed expression in the mutant larvae. FBgn0010551 is the fly homologue of prohibitin (PHB) 2, a nuclear encoded mitochondrial protein conserved throughout eukaryotes. Together with PHB1, it forms the PHB complex, which has been suggested to function as a membrane-associated chaperone/holdase, stabilizing unfolded proteins in the mitochondrial inner membrane (Nijtmans *et al.* 2000, reviewed by Artal-Sanz and Tavernarakis 2009). Disruption of the PHB complex has been found to cause mitochondrial defects in yeast. In multicellular organisms, the PHB complex is essential for development, and prohibitin depletion causes disruption of mitochondrial morphology in *C. elegans* (Artal-Sanz *et al.* 2003, Van Aken *et al.* 2007, Merkwirth and Langer 2009). Consistent with its important role, FBgn0010551 is an essential gene in *Drosophila* (Spradling *et al.* 1999).

Gene ontology term analysis on the array results pointed towards effects on pathways related to nucleotide, amino acid and fatty acid metabolism as well as a number of mitochondrial genes. A separate analysis on the expression of mitochondrial genes revealed progressive changes towards the 3rd instar. To obtain further evidence on a possible mitochondrial defect, we compared our data to the microarray study by Fernández-Ayala *et al.* (2010). They investigated a fly line with a mutation in *technical knockout* (*tko*) gene, which encodes an essential mitochondrial protein. We did these comparisons to seek for similarities between our data and a mitochondrial mutant *Drosophila*, because *prohibitin* mutants have not been screened on the whole-genome level in *Drosophila*. The comparisons revealed many similarities in the two sets of data: genes related to nucleotide, amino acid and fatty acid metabolism were implicated in both datasets. 15% of the significantly changed genes in our analysis were changed also in the *tko* mutant, and approximately 70% of those showed change in the same direction, which constituted a statistically highly significant overlap. Considering that Fernández-Ayala *et al.* (2010) used adult flies in their analysis, the similarities between the two datasets are remarkable. These results suggest that mitochondrial dysfunction has a significant role in the lethality caused by the U6atac mutation and strengthen the hypothesis that the *Drosophila prohibitin* gene homologue plays a central part in the mutation phenotype.

The surprising finding that genes with metabolic functions are implicated in the U12-type splicing mutant suggests that U12-type splicing efficiency can have an effect on fundamental metabolic pathways. While it is likely that other U12-type intron-containing genes contribute to the U6atac mutant phenotype, the PHB2 homologue FBgn0010551 seems a good candidate to explain many of the observed effects.

4.3 U12-dependent spliceosome functions in the nucleus (II)

4.3.1 The snRNA components of the U12-type spliceosome are localized in the nucleus

Since the discovery of a second spliceosome, theories have been sought to explain the need for two spliceosomal systems in complex eukaryotes. Recently, König *et al.* (2007) presented a provocative hypothesis that the U12-type spliceosome functions in the cytoplasm. Their study was in contradiction with earlier findings that have localized U11 and U12 snRNAs in the nucleus of human cells (Matera and Ward 1993) and U11 in the nucleus of Drosophila cells (Schneider *et al.* 2004). In addition, GFP-tagged U12-type spliceosome-specific proteins U11-35K, U11/U12-31K (also known as ZCRB1, MADP-1) and U11/U12-65K (RNPC3) are also predominantly nuclear in Arabidopsis and human cells (Zhao *et al.* 2003, Lorković *et al.* 2005, Wang *et al.* 2007). The contradictory previous studies were not addressed in the study by König *et al.* (2007).

We set out to clarify the matter by investigating the localization of all snRNA and protein components specific to the U12-type spliceosome. First, we studied the localization of spliceosomal snRNAs in mouse tissues. The snRNAs U1, U2, U6, U11, U12, U4atac and U6atac were detected in the nuclei of different mouse tissues by nonradioactive *in situ* hybridization on tissue sections using digoxigenin-labelled full-length RNA probes. In each case, the probe signal was overlapping with nuclear Hoechst counterstaining. No differences were observed in the localization of major and minor spliceosome snRNAs.

To confirm the results with a different detection method, the localization of U2, U12, U4atac and U6atac snRNAs was studied in mouse brain sections by radioactive *in situ* hybridization. Also in this case, the signal from snRNAs was found overlapping with nuclear counterstaining.

Next, we compared the subnuclear localization of U11 and U12 snRNAs to that of U2 and U4 in human HeLa cells by fluorescent *in situ* hybridization. Nuclear staining with several more intensely stained foci, likely Cajal bodies, was observed. Again, the major and minor snRNAs tested showed similar distribution. Thus, *in situ* hybridizations with three different detection methods on mouse tissues and human cells indicated that spliceosomal snRNAs are nuclear.

4.3.2 Protein and snRNA components of the U12-dependent spliceosome are detected primarily in nuclear fractions of cells

We also tested whether cellular fractionation would corroborate the results obtained by *in situ* hybridization. We prepared nuclear and cytoplasmic fractions from HeLa cells, isolated RNA and proteins from parallel fractions and detected spliceosomal components by Northern and Western blotting. Like U1, U2, U4 and U5 of the major spliceosome, the minor spliceosome snRNAs U11, U12 and U4atac were strongly enriched in nuclear fractions. 5S RNA and tRNA were detected predominantly in cytoplasmic fractions, indicating that the

fractions were well separated. Both U6 and U6atac appeared in nuclear and cytoplasmic extracts in approximately equal amounts, which is surprising but consistent with earlier reports (Fury and Zieve 1996). In the light of these results, the choice of König *et al.* (2007) to compare the distribution of U6atac to that of U2 in cellular fractions seems unfortunate.

Western blot analysis indicated that the U11-specific proteins 59K, 35K and 25K are predominantly found in the nuclear fraction, as are the snRNP-associated U1-70K, U5-15K, U4/U6-61K, SF3a66, SF3b155 and the splicing factor SF1. In contrast, Lsm1, a component of the cytoplasmic Lsm1-7 complex, was detected mainly in the cytoplasmic fraction. Taken together, the components of the U12-type spliceosome are predominantly nuclear and do not significantly differ in their localization from their counterparts in the U2-type spliceosome.

In addition to localization data, also indirect evidence points to nuclear localization of the U12-type spliceosome. U12-type splicing is predominantly cotranscriptional (Singh and Padgett 2009). Additionally, communication between U2-type and U12-type spliceosomes has been observed in exon definition interactions and in the recruitment of the major spliceosome by U11 at the USSE element, which controls the levels of 48K and 65K proteins (Wu and Krainer 1996, Verbeeren *et al.* 2010). Furthermore, inhibition of U12-type splicing often leads to activation of cryptic U2-type splice sites (Tarn and Steitz 1996b, Turunen *et al.* 2008), which would have devastating effects on the expression of genes containing U12-type introns *in vivo*. This suggests that some minor spliceosome-specific factor(s), most likely components of the U11/U12 di-snRNP, must be present at the transcription site to bind the U12-type intron and prevent the major spliceosome from interfering.

4.3.3 The nuclear dogma still holding

What could explain the disagreement between our results and the study of König *et al.* (2007)? They used locked nucleic acid oligonucleotide probes to detect snRNAs by *in situ* hybridization in zebrafish tissues and mouse fibroblasts. Each oligonucleotide contained only one digoxigenin molecule. This may not be sufficient to provide enough sensitivity to accurately detect the low-abundance minor snRNAs. König *et al.* (2007) also detected U6atac in cytoplasmic extracts but, as noted above, the presence of U6atac (and U6) in cytoplasmic fractions does not warrant the claim that splicing should occur outside the nucleus. The Lsm2-8 proteins bound by both U6 and U6atac function as nuclear retention/localization signals, and U6 injected into *Xenopus* oocyte cytoplasm is imported into the nucleus (Hamm and Mattaj 1989, Spiller *et al.* 2007). No U6atac-specific proteins have been found, which makes the existence of a mechanism resulting in a different subcellular localization for U6 and U6atac seem unlikely. In general, the numerous similarities in the structure, biogenesis and function between U2- and U12-type spliceosomes give no reason to believe that the two systems would differ in their subcellular localization.

As a response to criticism towards their paper König *et al.* (2007), König and Müller (2008) raised several arguments against the reliability of our results. First, they argued that their data were obtained in a different system (zebrafish tissues). Yet, in their report (König *et al.* 2007) they extrapolated their results on a general level by offering the segregation of the U12-type spliceosome in the cytoplasm as an explanation for the existence of two separate spliceosomes. The range of distantly related organisms (vertebrates, *Drosophila*, *Arabidopsis*; see **4.3.1**) in which U12-type spliceosome components have been found to reside in the

nucleus demonstrates that exceptions to the rule, if any existed, would be highly uncommon. Second, König and Müller (2008) questioned the specificity of the probes used in our *in situ* hybridizations since we did not include controls for probe specificity in fluorescent *in situ* hybridizations done on HeLa cells. Nevertheless, we did show control hybridizations on mouse tissue sections with sense-strand probes for U2 and U12, which did not give any detectable signal. König *et al.* (2007) used mismatch probes to demonstrate specificity of their oligonucleotide probes. However, control probes cannot be used to prove the specificity of cognate probes. They can only demonstrate that hybridization conditions do not introduce false signals with any probe. In our study, we used Northern blotting to verify the specificity of our probes (Supporting information in **II**). Third, König and Müller (2008) pointed out the difficulty of achieving a complete separation of nuclear and cytoplasmic fractions. The same criticism can be directed towards their own fractionation experiments, in which the presence of U6atac in cytoplasmic fractions was taken as evidence on the localization of the U12-type spliceosome. Considering the distribution of U6 in cellular fractions (see **4.3.2**), their conclusions were not substantiated. Fourth, König and Müller (2008) suggested that the DNA counterstain used in our *in situ* hybridizations would cause problems to specificity. This suspicion seems unwarranted since Hoechst staining is a commonly used procedure that has not been reported to cause problems with *in situ* techniques.

Later, there have been other reports determining U12-type splicing as a nuclear event in *Xenopus laevis* oocytes (Friend *et al.* 2008) and human cells (Singh and Padgett 2009). In conclusion, the nuclear localization of the U12-type spliceosome seems firmly established.

5 Concluding remarks

The existence of two spliceosomal systems in complex eukaryotes has evoked many questions concerning their differences and the rationale for maintaining two distinct spliceosomes. There is considerable evidence that U12-type introns limit the expression of their host genes. This additional means of gene expression control raises important questions: how is the activity of the U12-type spliceosome controlled, and which pathways and functions are affected by the level of splicing efficiency?

The low abundance of U12-type spliceosome components has been suggested to limit its activity. Earlier *in vitro* studies have shown that the slow phase of spliceosome assembly occurs after intron recognition, before catalytic activation; therefore this study was concentrated on investigating the components limiting the amount of the U4atac/U6atac.U5 tri-snRNP. The results suggested that the levels of U4atac snRNA indeed limit the formation of the minor tri-snRNP, but no effects on the expression of endogenous genes were detected. If one or some spliceosomal components are limiting spliceosome activity *in vivo*, they probably are components of the U11/U12 intron recognition complex. On the other hand, in cells containing normal levels of U12-type spliceosome components, the spliceosome does not appear to be saturated even by overexpression of ectopic U12-type introns. It is possible that earlier studies on *in vitro* kinetics do not accurately correlate with the situation in living cells where splicing occurs cotranscriptionally.

The second question about the transcriptome-level effects of low U12-type splicing efficiency was studied here using *Drosophila*. Despite their small number, U12-type introns in the fly exhibit the same preference to reside in genes involved in 'information processing' functions as mammalian U12-type introns. These pathways have been thought to be the main targets for regulation by U12-type splicing activity. This work suggests that regulation by splicing efficiency may affect an entirely different category of genes with fundamental metabolic functions. In the case of *Drosophila* larvae, the effects of the U12-type splicing defect could potentially be largely attributed to a single gene, the fly homologue of mitochondrial PHB2. The genes targeted by the rate-limiting splicing process may be different in different tissues, developmental stages or environmental situations. The U12-type intron in PHB2 gene is conserved from insects to vertebrates including human. It would be interesting to know whether the genes primarily affected by changes in U12-type splicing efficiency in mammals are similar in function to those implicated in splicing mutant *Drosophila*. Studying the primary target genes and pathways in different organisms could help decipher the role of the U12-type spliceosome in gene expression.

The intriguing question of why eukaryotes maintain two spliceosomal systems has prompted many theories since the discovery of the U12-type spliceosome. Unfortunately, some have not been backed up by solid evidence, such as the idea that the U12-type spliceosome would function in the cytoplasm. In this study, the localization of U12-type spliceosome components was comprehensively investigated and found to support the dogma of nuclear splicing. In the light of this work and other recent publications, the question of the subcellular localization of the U12-type spliceosome seems settled.

Acknowledgments

This study has been carried out at the Institute of Biotechnology, University of Helsinki. I thank Viikki Graduate School in Biosciences, Alfred Kordelin Foundation, and University of Helsinki Research Funds for financial support.

I am grateful to my supervisor Mikko Frilander for all his support during the years we have worked together, for always having time to give advice and for seeing promising results where I have seen only smudge. I also want to give warm thanks to all members of the RNA splicing lab for creating such a nice working atmosphere: Lilli, Elina, Janne, Jens, Jouni and Annukka. I especially want to thank Lilli and Annukka for all the help with experiments and Jouni for help with making the most of the data in my latest article.

I thank professor Marie Öhman for accepting the invitation to be my opponent, and professor Tapio Palva for accepting to act as custos at my defense. I also thank Marjo Salminen and professor Ykä Helariutta for reviewing my thesis and giving insightful comments and suggestions. Marjo was also in my follow-up group together with professor Pekka Lappalainen. I thank them for their advice and ideas during my graduate work.

In the *Drosophila* project, we had a pleasant collaboration with Petri Auvinen and Tapio Heino. I thank Petri and Dario Greco for sharing their expertise on microarrays. I also thank Tapio and the members of his group, Mari, Riitta and Vassileios, for their advice and help on fly work and for having me and my flies around occupying space in the fly facilities. Thanks also to all other collaborators home and abroad.

I thank professor Harri Savilahti, whose group were our long-time lab neighbours, and the former members of the Savilahti laboratory, Maria, Anja, Tiina, Heikki, Hilkka, Eini, Pirjo, Sari and others, for contributing to the pleasant working environment and also for their advice with experiments and lending of equipment.

For most of my time in BI, our group has belonged to the Developmental Biology Research Programme. I want to thank all the Development people and especially the Director of the programme, professor Irma Thesleff, for giving me a glance into a world different from that of RNA splicing, and of course for the refreshing retreats in Tvärminne and Hyytiälä. I also thank the former Director of the Institute, Mart Saarma, and his successor, Tomi Mäkelä, for providing an inspiring working environment.

I thank Viikki Graduate School in Biosciences for organizing many interesting courses based on students' requests and its former coordinator Eeva Sievi for prompt answers to all tricky questions concerning studies, university bureaucracy and everything possible.

I give my heartfelt thanks to my best friends since our first year at the University, Elina and Eija. Elina, thanks for all those coffee breaks back when we were both graduate students, and for all your advice, some of which I actually took. I thank you both for always being there.

Kiitokset äidille ja isälle sekä henkisestä että materiaalisesta tuesta vuosien varrella. Kiitos että olette aina kannustaneet ja uskoneet minuun, ja myös siitä että olette pitäneet huolta karvaisesta kaveristani Piipulista.

Helsinki, September 2010

Heli

References

Abelson, J., Trotta, C.R. and Li, H. (1998) tRNA splicing. *J. Biol. Chem.* 273: 12685-12688.

Abovich, N. and Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 89: 403-412.

Achsel, T., Brahms, H., Kastner, B., Bachi, A., Wilm, M. and Lührmann, R. (1999) A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *EMBO J.* 18: 5789-5802.

Alioto, T.S. (2007) U12DB: a database of orthologous U12-type spliceosomal introns. *Nucleic Acids Res.* 35: D110-115.

Artal-Sanz, M., Tsang, W.Y., Willems, E.M., Grivell, L.A., Lemire, B.D., van der Spek, H. and Nijtmans, L.G.J. (2003) The mitochondrial prohibitin complex is essential for embryonic viability and germline function in Caenorhabditis elegans. *J. Biol. Chem.* 278: 32091-32099.

Artal-Sanz, M. and Tavernarakis, N. (2009) Prohibitin and mitochondrial biology. *Trends Endocrinol. Metab.* 20: 394-401.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J. and Frey, B.J. (2010) Deciphering the splicing code. *Nature* 465: 53-59.

Basu, M.K., Makalowski, W., Rogozin, I.B. and Koonin, E.V. (2008) U12 intron positions are more strongly conserved between animals and plants than U2 intron positions. *Biol. Direct* 3: 19.

Behzadnia, N., Hartmuth, K., Will, C.L. and Lührmann, R. (2006) Functional spliceosomal A complexes can be assembled in vitro in the absence of a penta-snRNP. *RNA* 12: 1738-1746.

Bell, M., Schreiner, S., Damianov, A., Reddy, R. and Bindereif, A. (2002) p110, a novel human U6 snRNP protein and U4/U6 snRNP recycling factor. *EMBO J.* 21: 2724-2735.

Benecke, H., Lührmann, R. and Will, C.L. (2005) The U11/U12 snRNP 65K protein acts as a molecular bridge, binding the U12 snRNA and U11-59K protein. *EMBO J.* 24: 3057-3069.

Berglund, J.A., Chua, K., Abovich, N., Reed, R. and Rosbash, M. (1997) The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 89: 781-787.

Beyer, A.L. and Osheim, Y.N. (1988) Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2: 754-765.

Bird, G., Zorio, D.A.R. and Bentley, D.L. (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol. Cell. Biol.* 24: 8963-8969.

Black, D.L., Chabot, B. and Steitz, J.A. (1985) U2 as well as U1 small nuclear ribonucleoproteins are involved in premessenger RNA splicing. *Cell* 42: 737-750.

Black, D.L. and Steitz, J.A. (1986) Pre-mRNA splicing in vitro requires intact U4/U6 small nuclear ribonucleoprotein. *Cell* 46: 697-704.

Bonen, L. (1993) Trans-splicing of pre-mRNA in plants, animals, and protists. *FASEB J.* 7: 40-46.

Bonen, L. and Vogel, J. (2001) The ins and outs of group II introns. *Trends Genet.* 17: 322-331.

Borah, S., Wong, A.C. and Steitz, J.A. (2009) Drosophila hnRNP A1 homologs Hrp36/Hrp38 enhance U2-type versus U12-type splicing to regulate alternative splicing of the prospero twintron. *Proc. Natl. Acad. Sci. U.S.A.* 106: 2577-2582.

Branlant, C., Krol, A., Ebel, J.P., Lazar, E., Haendler, B. and Jacob, M. (1982) U2 RNA shares a structural domain with U1, U4 and U5 RNAs. *EMBO J.* 1: 1259-1265.

Bringmann, P., Appel, B., Rinke, J., Reuter, R., Theissen, H. and Lührmann, R. (1984) Evidence for the existence of snRNAs U4 and U6 in a single ribonucleoprotein complex and for their association by intermolecular base pairing. *EMBO J.* 3: 1357-1363.

Bringmann, P. and Lührmann, R. (1986) Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *EMBO J.* 5: 3509-3516.

Brock, J.E., Dietrich, R.C. and Padgett, R.A. (2008) Mutational analysis of the U12-dependent branch site consensus sequence. *RNA* 14: 2430-2439.

Bühler, D., Raker, V., Lührmann, R. and Fischer, U. (1999) Essential role for the tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum. Mol. Genet.* 8: 2351-2357.

Burge, C.B., Padgett, R.A. and Sharp, P.A. (1998) Evolutionary fates and origins of U12-type introns. *Mol. Cell* 2: 773-785.

Burge, C.B., Tuschl, T. and Sharp, P.A. Splicing of precursors to mRNAs by the spliceosome in *The RNA world* (eds. Gesteland, R.F., Cech, T.R. and Atkins, J.F.) 2nd edition, pp. 525-560 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999).

Burghes, A.H.M. and Beattie, C.E. (2009) Spinal muscular atrophy: why do low levels of survival motor neuron protein make motor neurons sick?. *Nat. Rev. Neurosci.* 10: 597-609.

Burnette, J.M., Miyamoto-Sato, E., Schaub, M.A., Conklin, J. and Lopez, A.J. (2005) Subdivision of large introns in Drosophila by recursive splicing at nonexonic elements. *Genetics* 170: 661-674.

Cascino, I., Fiucci, G., Papoff, G. and Ruberti, G. (1995) Three functional soluble forms of the human apoptosis-inducing Fas molecule are produced by alternative splicing. *J. Immunol.* 154: 2706-2713.

Castle, J.C., Zhang, C., Shah, J.K., Kulkarni, A.V., Kalsotra, A., Cooper, T.A. and Johnson, J.M. (2008) Expression of 24,426 human alternative splicing events and predicted cis regulation in 48 tissues and cell lines. *Nat. Genet.* 40: 1416-1425.

Cech, T.R. (1986) The generality of self-splicing RNA: relationship to nuclear mRNA splicing. *Cell* 44: 207-210.

Chabot, B., Black, D.L., LeMaster, D.M. and Steitz, J.A. (1985) The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. *Science* 230: 1344-1349.

Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel, R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T., Rosenberg, T., Webster, A.R., Bird, A.C., Gal, A., Hunt, D., Vithana, E.N. and Bhattacharya, S.S. (2002) Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.* 11: 87-92.

Chalfant, C.E., Watson, J.E., Bisnauth, L.D., Kang, J.B., Patel, N., Obeid, L.M., Eichler, D.C. and Cooper, D.R. (1998) Insulin regulates protein kinase CbetaII expression through enhanced exon inclusion in L6 skeletal muscle cells. A novel mechanism of insulin- and insulin-like growth factor-i-induced 5' splice site selection. *J. Biol. Chem.* 273: 910-916.

Chang, D.D. and Sharp, P.A. (1989) Regulation by HIV Rev depends upon recognition of splice sites. *Cell* 59: 789-795.

Chen, W., Luo, L. and Zhang, L. (2010) The organization of nucleosomes around splice sites. *Nucleic Acids Res.* 38: 2788-2798.

Cheng, S.C. and Abelson, J. (1987) Spliceosome assembly in yeast. *Genes Dev.* 1: 1014-1027.

Conti, E. and Izaurralde, E. (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* 17: 316-325.

Crick, F. (1979) Split genes and RNA splicing. *Science* 204: 264-271.

Crooks, G.E., Hon, G., Chandonia, J.M. and Brenner, S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.* 14: 1188-1190.

Custódio, N., Carmo-Fonseca, M., Geraghty, F., Pereira, H.S., Grosveld, F. and Antoniou, M. (1999) Inefficient processing impairs release of RNA from the site of transcription. *EMBO J.* 18: 2855-2866.

Damianov, A., Schreiner, S. and Bindereif, A. (2004) Recycling of the U12-type spliceosome requires p110, a component of the U6atac snRNP. *Mol. Cell. Biol.* 24: 1700-1708.

Darnell, J.E.J. (1978) Implications of RNA-RNA splicing in evolution of eukaryotic cells. *Science* 202: 1257-1260.

Darreh-Shori, T., Hellström-Lindahl, E., Flores-Flores, C., Guan, Z.Z., Soreq, H. and Nordberg, A. (2004) Long-lasting acetylcholinesterase splice variations in anticholinesterase-treated Alzheimer's disease patients. *J. Neurochem.* 88: 1102-1113.

Das, R., Dufu, K., Romney, B., Feldt, M., Elenko, M. and Reed, R. (2006) Functional coupling of RNAP II transcription to spliceosome assembly. *Genes Dev.* 20: 1100-1109.

Dávila López, M., Rosenblad, M.A. and Samuelsson, T. (2008) Computational screen for spliceosomal RNA genes aids in defining the phylogenetic distribution of major and minor spliceosomal components. *Nucleic Acids Res.* 36: 3001-3010.

Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H. and Lührmann, R. (2006) protein composition and electron microscopy structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol. Cell. Biol.* 26: 5528-5543.

Denis, M.M., Tolley, N.D., Bunting, M., Schwertz, H., Jiang, H., Lindemann, S., Yost, C.C., Rubner, F.J., Albertine, K.H. and Swoboda, K.J. (2005) Escaping the nuclear confines: Signal-dependent pre-mRNA splicing in anucleate platelets. *Cell* 122: 379-391.

Dietrich, R.C., Incorvaia, R. and Padgett, R.A. (1997) Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. *Mol. Cell* 1: 151-160.

Dietrich, R.C., Shukla, G.C., Fuller, J.D. and Padgett, R.A. (2001) Alternative splicing of U12-dependent introns in vivo responds to purine-rich enhancers. *RNA* 7: 1378-1388.

Domitrovich, A.M. and Kunkel, G.R. (2003) Multiple, dispersed human U6 small nuclear RNA genes with varied transcriptional efficiencies. *Nucleic Acids Res.* 31: 2344-2352.

Doolittle, W.F. (1978) Genes in pieces: were they ever together?. *Nature* 272: 581-582.

Dorn, R., Reuter, G. and Loewendorf, A. (2001) Transgene analysis proves mRNA trans-splicing at the complex mod(mdg4) locus in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* 98: 9724-9729.

Dutertre, M., Lacroix-Triki, M., Driouch, K., de la Grange, P., Gratadou, L., Beck, S., Millevoi, S., Tazi, J., Lidereau, R., Vagner, S. and Auboeuf, D. (2010) Exon-based clustering of murine breast tumor transcriptomes reveals alternative exons whose expression is associated with metastasis. *Cancer Res.* 70: 896-905.

Fabrizio, P., Laggerbauer, B., Lauber, J., Lane, W.S. and Lührmann, R. (1997) An evolutionarily conserved U5 snRNP-specific protein is a GTP-binding factor closely related to the ribosomal translocase EF-2. *EMBO J.* 16: 4092-4106.

Fairbrother, W.G., Yeh, R., Sharp, P.A. and Burge, C.B. (2002) Predictive identification of exonic splicing enhancers in human genes. *Science* 297: 1007-1013.

Faustino, N.A. and Cooper, T.A. (2003) Pre-mRNA splicing and human disease. *Genes Dev.* 17: 419-437.

Fernández-Ayala, D.J.M., Chen, S., Kemppainen, E., O'Dell, K.M.C. and Jacobs, H.T. (2010) Gene expression in a Drosophila model of mitochondrial disease. *PLoS ONE* 5: e8549.

Fischer, U. and Lührmann, R. (1990) An essential signaling role for the m3G cap in the transport of U1 snRNP to the nucleus. *Science* 249: 786-790.

Fischer, U., Sumpter, V., Sekine, M., Satoh, T. and Lührmann, R. (1993) Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that binds a transport receptor independently of the m3G cap. *EMBO J.* 12: 573-583.

Fischer, U., Liu, Q. and Dreyfuss, G. (1997) The SMN-SPI1 complex has essential role in spliceosomal snRNP biogenesis. *Cell* 90: 1023-1029.

Fong, Y.W. and Zhou, Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. *Nature* 414: 929-933.

Forbes, D.J., Kirschner, M.W., Caput, D., Dahlberg, J.E. and Lund, E. (1984) Differential expression of multiple U1 small nuclear RNAs in oocytes and embryos of Xenopus laevis. *Cell* 38: 681-689.

Fouser, L.A. and Friesen, J.D. (1987) Effects on mRNA splicing of mutations in the 3' region of the Saccharomyces cerevisiae actin intron. *Mol. Cell. Biol.* 7: 225-230.

Friend, K., Kolev, N.G., Shu, M. and Steitz, J.A. (2008) Minor-class splicing occurs in the nucleus of the Xenopus oocyte. *RNA* 14: 1459-1462.

Frilander, M.J. and Steitz, J.A. (1999) Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. *Genes Dev.* 13: 851-863.

Frilander, M.J. and Steitz, J.A. (2001) Dynamic exchanges of RNA interactions leading to catalytic core formation in the U12-dependent spliceosome. *Mol. Cell* 7: 217-226.

Fury, M.G. and Zieve, G.W. (1996) U6 snRNA maturation and stability. *Exp. Cell Res.* 228: 160-163.

Gabanella, F., Butchbach, M.E.R., Saieva, L., Carissimi, C., Burghes, A.H.M. and Pellizzoni, L. (2007) Ribonucleoprotein assembly defects correlate with spinal muscular atrophy severity and preferentially affect a subset of spliceosomal snRNPs. *PLoS ONE* 2: e921.

García-Blanco, M.A., Jamison, S.F. and Sharp, P.A. (1989) Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev.* 3: 1874-1886.

García-Blanco, M.A., Baraniak, A.P. and Lasda, E.L. (2004) Alternative splicing in disease and therapy. *Nat. Biotechnol.* 22: 535-546.

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S. and Biamonti, G. (2005) Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. *Mol. Cell* 20: 881-890.

Gilbert, W. (1978) Why genes in pieces?. *Nature* 271: 501.

Gilbert, W. (1987) The exon theory of genes. *Cold Spring Harb. Symp. Quant. Biol.* 52: 901-905.

Glanzer, J., Miyashiro, K.Y., Sul, J.Y., Barrett, L., Belt, B., Haydon, P. and Eberwine, J. (2005) RNA splicing capability of live neuronal dendrites. *Proc. Natl. Acad. Sci. U.S.A.* 102: 16859-16864.

Graveley, B.R. (2000) Sorting out the complexity of SR protein functions. *RNA* 6: 1197-1211.

Graveley, B.R. (2001) Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17: 100-107.

Hall, S.L. and Padgett, R.A. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice- sites. *J. Mol. Biol.* 239: 357-365.

Hall, S.L. and Padgett, R.A. (1996) Requirement of U12 snRNA for in vivo splicing of a minor class of eukaryotic nuclear pre- mRNA introns. *Science* 271: 1716-1718.

Hallegger, M., Llorian, M. and Smith, C.W.J. (2010) Alternative splicing: global insights. *FEBS J.* 277: 856-866.

Hamer, D.H. and Leder, P. (1979) Splicing and the formation of stable RNA. *Cell* 18: 1299-1302.

Hamm, J. and Mattaj, I.W. (1989) An abundant U6 snRNP found in germ cells and embryos of Xenopus laevis. *EMBO J.* 8: 4179-4187.

Hamm, J. and Mattaj, I.W. (1990) Monomethylated cap structures facilitate RNA export from the nucleus. *Cell* 63: 109-118.

Han, K., Yeo, G., An, P., Burge, C.B. and Grabowski, P.J. (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.* 3: e158.

Hanley, B.A. and Schuler, M.A. (1991) Developmental expression of plant snRNAs. *Nucleic Acids Res.* 19: 6319-6325.

Hartmuth, K., Urlaub, H., Vornlocher, H., Will, C.L., Gentzel, M., Wilm, M. and Lührmann, R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc. Natl. Acad. Sci. U.S.A.* 99: 16719-16724.

Hashimoto, C. and Steitz, J.A. (1984) U4 and U6 RNAs coexist in a single small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* 12: 3283-3293.

Hastings, M. and Krainer, A. (2001) Functions of SR proteins in the U12-dependent AT-AC pre-mRNA splicing pathway. *RNA* 7: 471-482.

Hatton, A.R., Subramaniam, V. and Lopez, A.J. (1998) Generation of alternative Ultrabithorax isoforms and stepwise removal of a large intron by resplicing at exon-exon junctions. *Mol. Cell* 2: 787-796.

Hausner, T.P., Giglio, L.M. and Weiner, A.M. (1990) Evidence for base-pairing between mammalian U2 and U6 small nuclear ribonucleoprotein particles. *Genes Dev.* 4: 2146-2156.

Herold, N., Will, C.L., Wolf, E., Kastner, B., Urlaub, H. and Lührmann, R. (2009) Conservation of the protein composition and electron microscopy structure of Drosophila melanogaster and human spliceosomal complexes. *Mol. Cell. Biol.* 29: 281-301.

Hirose, T., Shu, M. and Steitz, J.A. (2004) Splicing of U12-type introns deposits an exon junction complex competent to induce nonsense-mediated mRNA decay. *Proc. Natl. Acad. Sci. U.S.A.* 101: 17976-17981.

Hodnett, J.L. and Busch, H. (1968) Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. *J. Biol. Chem.* 243: 6334-6342.

Hoffman, B.E. and Grabowski, P.J. (1992) U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev.* 6: 2554-2568.

Huber, J., Cronshagen, U., Kadokura, M., Marshallsay, C., Wada, T., Sekine, M. and Lührmann, R. (1998) Snurportin1, an m3G-cap-specific nuclear import receptor with a novel domain structure. *EMBO J.* 17: 4114-4126.

Incorvaia, R. and Padgett, R.A. (1998) Base pairing with U6atac snRNA is required for 5' splice site activation of U12-dependent introns in vivo. *RNA* 4: 709-718.

Jackson, I.J. (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.* 19: 3795-3798.

Jády, B.E., Darzacq, X., Tucker, K.E., Matera, A.G., Bertrand, E. and Kiss, T. (2003) Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *EMBO J.* 22: 1878-1888.

Jarrell, K.A., Dietrich, R.C. and Perlman, P.S. (1988) Group II intron domain 5 facilitates a trans-splicing reaction. *Mol. Cell. Biol.* 8: 2361-2366.

Jeffares, D.C., Mourier, T. and Penny, D. (2006) The biology of intron gain and loss. *Trends Genet.* 22: 16-22.

Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y. and Darnell, R.B. (2000) Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron* 25: 359-371.

Jin, Y., Suzuki, H., Maegawa, S., Endo, H., Sugano, S., Hashimoto, K., Yasuda, K. and Inoue, K. (2003) A vertebrate RNA-binding protein Fox-1 regulates tissue-specific splicing via the pentanucleotide GCAUG. *EMBO J.* 22: 905-912.

Johnson, J.M., Castle, J., Garrett-Engele, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R. and Shoemaker, D.D. (2003) Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302: 2141-2144.

Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N. and Moore, M.J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *RNA* 8: 426-439.

Jurica, M.S. and Moore, M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* 12: 5-14.

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D. and Krainer, A.R. (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nat. Struct. Mol. Biol.* 14: 185-193.

Kataoka, N., Yong, J., Kim, V.N., Velazquez, F., Perkinson, R.A., Wang, F. and Dreyfuss, G. (2000) Pre-mRNA splicing imprints mRNA in the nucleus with a novel RNA-binding protein that persists in the cytoplasm. *Mol. Cell* 6: 673-682.

Kolossova, I. and Padgett, R.A. (1997) U11 snRNA interacts in vivo with the 5' splice site of U12-dependent (AU-AC) pre-mRNA introns. *RNA* 3: 227-233.

Konarska, M.M. and Sharp, P.A. (1987) Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* 49: 763-774.

König, H., Matter, N., Bader, R., Thiele, W. and Müller, F. (2007) Splicing segregation: the minor spliceosome acts outside the nucleus and controls cell proliferation. *Cell* 131: 718-729.

König, H. and Müller, F. (2008) Minor splicing: Nuclear dogma still in question. *Proc. Natl. Acad. Sci. U.S.A.* 105: E37.

Koonin, E.V. (2006) The origin of introns and their role in eukaryogenesis: a compromise solution to the introns-early versus introns-late debate?. *Biol. Direct* 1: 22.

Kotovic, K.M., Lockshon, D., Boric, L. and Neugebauer, K.M. (2003) Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol. Cell. Biol.* 23: 5768-5779.

Krainer, A.R. and Maniatis, T. (1985) Multiple factors includung the small nuclear ribonucleoproteins U1 and U2 are necessary for pre-mRNA splicing in vitro. *Cell* 42: 725-736.

Krämer, A., Keller, W., Appel, B. and Lührmann, R. (1984) The 5' terminus of the RNA moiety of U1 small nuclear ribonucleoprotein particles is required for the splicing of messenger RNA precursors. *Cell* 38: 299-307.

Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* 31: 147-157.

Kunkel, G.R., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) U6 small nuclear RNA is transcribed by RNA polymerase III. *Proc. Natl. Acad. Sci. U.S.A.* 83: 8575-8579.

Kupfer, D.M., Drabenstot, S.D., Buchanan, K.L., Lai, H., Zhu, H., Dyer, D.W., Roe, B.A. and Murphy, J.W. (2004) Introns and splicing elements of five diverse fungi. *Eukaryotic Cell* 3: 1088-1100.

Laggerbauer, B., Achsel, T. and Lührmann, R. (1998) The human U5-200kD DEXH-box protein unwinds U4/U6 RNA duplices in vitro. *Proc. Natl. Acad. Sci. U.S.A.* 95: 4188-4192.

Lamond, A.I. and Spector, D.L. (2003) Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4: 605-612.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K. et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.

Le Hir, H., Moore, M.J. and Maquat, L.E. (2000) Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev.* 14: 1098-1108.

Lefebvre, S., Bürglen, L., Reboullet, S., Clermont, O., Burlet, P., Viollet, L., Benichou, B., Cruaud, C., Millasseau, P., Zeviani, M., Le Paslier, D., Frézal, J., Cohen, D., Weissenbach, J., Munnich, A. and Melki, J. (1995) Identification and characterization of a spinal muscular atrophy-determining gene. *Cell* 80: 155-165.

Lehmeier, T., Foulaki, K. and Lührmann, R. (1990) Evidence for three distinct D proteins, which react differentially with anti-Sm autoantibodies, in the cores of the major snRNAPs U1, U2, U4/U6 and U5. *Nucleic Acids Res.* 18: 6475-6484.

Lerner, M.R. and Steitz, J.A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl. Acad. Sci. U.S.A.* 76: 5495-5499.

Levine, A. and Durbin, R. (2001) A computational scan for U12-dependent introns in the human genome sequence. *Nucleic Acids Res.* 29: 4006-4013.

Lewis, B.P., Green, R.E. and Brenner, S.E. (2003) Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U.S.A.* 100: 189-192.

Lewis, J.D., Izaurralde, E., Jarmolowski, A., McGuigan, C. and Mattaj, I.W. (1996) A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev.* 10: 1683-1698.

Liautard, J.P., Sri-Widada, J., Brunel, C. and Jeanteur, P. (1982) Structural organization of ribonucleoproteins containing small nuclear RNAs from HeLa cells. Proteins interact closely with a similar structural domain of U1, U2, U4 and U5 small nuclear RNAs. *J. Mol. Biol.* 162: 623-643.

Lim, L.P. and Burge, C.B. (2001) A computational analysis of sequence features involved in recognition of short introns. *Proc. Natl. Acad. Sci. U.S.A.* 98: 11193-11198.

Lin, C., Mount, S.M., Jarmołowski, A. and Makałowski, W. (2010) Evolutionary dynamics of U12-type spliceosomal introns. *BMC Evol. Biol.* 10: 47.

Liu, F. and Gong, C. (2008) Tau exon 10 alternative splicing and tauopathies. *Mol Neurodegener* 3: 8.

López-Bigas, N., Audit, B., Ouzounis, C., Parra, G. and Guigó, R. (2005) Are splicing mutations the most frequent cause of hereditary disease?. *FEBS Lett.* 579: 1900-1903.

Lorković, Z.J., Lehner, R., Forstner, R. and Barta, A. (2005) Evolutionary conservation of minor U12-type spliceosome between plants and humans. *RNA* 11: 1095-1107.

Luberg, K., Wong, J., Weickert, C.S. and Timmusk, T. (2010) Human TrkB gene: novel alternative transcripts, protein isoforms and expression pattern in the prefrontal cerebral cortex during postnatal development. *J. Neurochem.* 113: 952-964.

Luco, R.F., Pan, Q., Tominaga, K., Blencowe, B.J., Pereira-Smith, O.M. and Misteli, T. (2010) Regulation of alternative splicing by histone modifications. *Science* 327: 996-1000.

Lund, E., Kahan, B. and Dahlberg, J.E. (1985) Differential control of U1 small nuclear RNA expression during mouse development. *Science* 229: 1271-1274.

Luo, H.R., Moreau, G.A., Levin, N. and Moore, M.J. (1999) The human Prp8 protein is a component of both U2- and U12-dependent spliceosomes. *RNA* 5: 893-908.

Madhani, H.D. and Guthrie, C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* 71: 803-817.

Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M. and Lührmann, R. (2002) Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298: 2205-2208.

Makarova, O.V., Makarov, E.M., Liu, S., Vornlocher, H. and Lührmann, R. (2002) Protein 61K, encoded by a gene (PRPF31) linked to autosomal dominant retinitis pigmentosa, is required for U4/U6·U5 tri-snRNP formation and pre-mRNA splicing. *EMBO J.* 21: 1148-1157.

Malca, H., Shomron, N. and Ast, G. (2003) The U1 snRNP base pairs with the 5' splice site within a penta-snRNP complex. *Mol. Cell. Biol.* 23: 3442-3455.

Mansilla, A., López-Sánchez, C., de la Rosa, E.J., García-Martínez, V., Martínez-Salas, E., de Pablo, F. and Hernández-Sánchez, C. (2005) Developmental regulation of a proinsulin messenger RNA generated by intron retention. *EMBO Rep.* 6: 1182–1187.

Martin, W. and Koonin, E.V. (2006) Introns and the origin of nucleus-cytosol compartmentalization. *Nature* 440: 41-45.

Marz, M., Kirsten, T. and Stadler, P.F. (2008) Evolution of spliceosomal snRNA genes in metazoan animals. *J. Mol. Evol.* 67: 594-607.

Maschhoff, K.L. and Padgett, R.A. (1993) The stereochemical course of the first step of pre-mRNA splicing. *Nucleic Acids Res.* 21: 5456-5462.

Matera, A.G. and Ward, D.C. (1993) Nucleoplasmic organization of small nuclear ribonucleoproteins in cultured human cells. *J. Cell Biol.* 121: 715-727.

Mattaj, I.W. (1986) Cap hypermethylation of U snRNA is cytoplasmic and dependent on snRNP protein binding. *Cell* 46: 905-911.

Mayeda, A. and Krainer, A.R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68: 365-375.

McKie, A.B., McHale, J.C., Keen, T.J., Tarttelin, E.E., Goliath, R., van Lith-Verhoeven, J.J., Greenberg, J., Ramesar, R.S., Hoyng, C.B., Cremers, F.P., Mackey, D.A., Bhattacharya, S.S., Bird, A.C., Markham, A.F. and Inglehearn, C.F. (2001) Mutations in the pre-mRNA splicing factor gene PRPC8 in autosomal dominant retinitis pigmentosa (RP13). *Hum. Mol. Genet.* 10: 1555-1562.

McNally, L.M., Yee, L. and McNally, M.T. (2006) Heterogeneous nuclear ribonucleoprotein H is required for optimal U11 small nuclear ribonucleoprotein binding to a retroviral RNA-processing control element: implications for U12-dependent RNA splicing. *J. Biol. Chem.* 281: 2478-2488.

Merendino, L., Guth, S., Bilbao, D., Martinez, C. and Valcárcel, J. (1999) Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature* 402: 838-841.

Merkwirth, C. and Langer, T. (2009) Prohibitin function within mitochondria: essential roles for cell proliferation and cristae morphogenesis. *Biochim. Biophys. Acta* 1793: 27-32.

Montzka, K.A. and Steitz, J.A. (1988) Additional low-abundance human small nuclear ribonucleoproteins: U11, U12 etc. *Proc. Natl. Acad. Sci. U.S.A.* 85: 8885-8889.

Moore, M.J. and Sharp, P.A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. *Nature* 365: 364-368.

Moore, M.J. and Proudfoot, N.J. (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* 136: 688-700.

Mount, S.M., Pettersson, I., Hinterberger, M., Karmas, A. and Steitz, J.A. (1983) The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell* 33: 509-518.

Mount, S.M. and Salz, H.K. (2000) Pre-messenger RNA Processing Factors in the Drosophila Genome. *J. Cell Biol.* 150: 37F-44.

Mount, S.M., Gotea, V., Lin, C., Hernandez, K. and Makalowski, W. (2007) Spliceosomal small nuclear RNA genes in 11 insect genomes. *RNA* 13: 5-14. **Narayanan, U., Ospina, J.K., Frey, M.R., Hebert, M.D. and Matera, A.G.** (2002) SMN, the spinal muscular atrophy protein, forms a pre-import snRNP complex with snurportin1 and importin beta. *Hum. Mol. Genet.* 11: 1785-1795.

Narayanan, U., Achsel, T., Lührmann, R. and Matera, A.G. (2004) Coupled in vitro import of U snRNPs and SMN, the spinal muscular atrophy protein. *Mol. Cell* 16: 223-234.

Newman, A.J. and Norman, C. (1992) U5 snRNA interacts with exon sequences at 5' and 3' splice sites. *Cell* 68: 743-754.

Nijtmans, L.G., de Jong, L., Artal Sanz, M., Coates, P.J., Berden, J.A., Back, J.W., Muijsers, A.O., van der Spek, H. and Grivell, L.A. (2000) Prohibitins act as a membrane-bound chaperone for the stabilization of mitochondrial proteins. *EMBO J.* 19: 2444-2451.

Niwa, M., Rose, S.D. and Berget, S.M. (1990) In vitro polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev.* 4: 1552-1559.

Niwa, M. and Berget, S.M. (1991) Mutation of the AAUAAA polyadenylation signal depresses in vitro splicing of proximal but not distal introns. *Genes Dev.* 5: 2086-2095.

Nottrott, S., Urlaub, H. and Lührmann, R. (2002) Hierarchical, clustered protein interactions with U4/U6 snRNA: a biochemical role for U4/U6 proteins. *EMBO J.* 21: 5527-5538.

Ohno, M., Segref, A., Bachi, A., Wilm, M. and Mattaj, I.W. (2000) PHAX, a mediator of U snRNA nuclear export whose activity is regulated by phosphorylation. *Cell* 101: 187-198.

Otake, L.R., Scamborova, P., Hashimoto, C. and Steitz, J.A. (2002) The divergent U12-type spliceosome is required for pre-mRNA splicing and is essential for development in Drosophila. *Mol. Cell* 9: 439-446.

Padgett, R.A., Konarska, M.M., Grabowski, P.J., Hardy, S.F. and Sharp, P.A. (1984) Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science* 225: 898-903.

Padgett, R.A. and Shukla, G.C. (2002) A revised model for U4atac/U6atac snRNA base pairing. *RNA* 8: 125-128.

Pagani, F., Raponi, M. and Baralle, F.E. (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl. Acad. Sci. U.S.A.* 102: 6368-6372.

Pan, Q., Shai, O., Misquitta, C., Zhang, W., Saltzman, A.L., Mohammad, N., Babak, T., Siu, H., Hughes, T.R., Morris, Q.D., Frey, B.J. and Blencowe, B.J. (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. *Mol. Cell* 16: 929-941.

Pannone, B.K., Xue, D. and Wolin, S.L. (1998) A role for the yeast La protein in U6 snRNP assembly: evidence that the La protein is a molecular chaperone for RNA polymerase III transcripts. *EMBO J.* 17: 7442-7453.

Park, E.J., Kim, J.H., Seong, R.H., Kim, C.G., Park, S.D. and Hong, S.H. (1999) Characterization of a novel mouse cDNA, ES18, involved in apoptotic cell death of T-cells. *Nucleic Acids Res.* 27: 1524-1530.

Parker, R., Siliciano, P.G. and Guthrie, C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell* 49: 229-239.

Patel, A.A., McCarthy, M. and Steitz, J.A. (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J.* 21: 3804-3815.

Patel, S.B. and Bellini, M. (2008) The assembly of a spliceosomal small nuclear ribonucleoprotein particle. *Nucleic Acids Res.* 36: 6482-6493.

Pellizzoni, L., Yong, J. and Dreyfuss, G. (2002) Essential role for the SMN complex in the specificity of snRNP assembly. *Science* 298: 1775-1779.

Pleiss, J.A., Whitworth, G.B., Bergkessel, M. and Guthrie, C. (2007) Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol. Cell* 27: 928-937.

Pozzoli, U., Sironi, M., Cagliani, R., Comi, G.P., Bardoni, A. and Bresolin, N. (2002) Comparative analysis of the human dystrophin and utrophin gene structures. *Genetics* 160: 793-798.

Query, C.C., Moore, M.J. and Sharp, P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.* 8: 587-597.

Raghunathan, P.L. and Guthrie, C. (1998a) A spliceosomal recycling factor that reanneals U4 and U6 small nuclear ribonucleoprotein particles. *Science* 279: 857-860.

Raghunathan, P.L. and Guthrie, C. (1998b) RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Current Biology* 8: 847-855.

Rappsilber, J., Ryder, U., Lamond, A.I. and Mann, M. (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res.* 12: 1231-1245.

Reddy, R., Ro-Choi, T.S., Henning, D. and Busch, H. (1974) Primary sequence of U-1 nuclear ribonucleic acid of Novikoff hepatoma ascites cells. *J. Biol. Chem.* 249: 6486-6494.

Reed, R. (2003) Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell Biol.* 15: 326-331.

Richard, H., Schulz, M.H., Sultan, M., Nürnberger, A., Schrinner, S., Balzereit, D., Dagand, E., Rasche, A., Lehrach, H., Vingron, M., Haas, S.A. and Yaspo, M. (2010) Prediction of alternative isoforms from exon expression levels in RNA-Seq experiments. *Nucleic Acids Res.* 38: e112.

Rinke, J. and Steitz, J.A. (1985) Association of the lupus antigen La with a subset of U6 snRNA molecules. *Nucleic Acids Res.* 13: 2617-2629.

Robberson, B.L., Cote, G.J. and Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* 10: 84-94.

Robertson, H.M. (1998) Two large families of chemoreceptor genes in the nematodes Caenorhabditis elegans and Caenorhabditis briggsae reveal extensive gene duplication, diversification, movement, and intron loss. *Genome Res.* 8: 449-463.

Romfo, C.M., Alvarez, C.J., van Heeckeren, W.J., Webb, C.J. and Wise, J.A. (2000) Evidence for splice site pairing via intron definition in Schizosaccharomyces pombe. *Mol. Cell. Biol.* 20: 7955-7970.

Roy, S.W. and Gilbert, W. (2005) Rates of intron loss and gain: implications for early eukaryotic evolution. *Proc. Natl. Acad. Sci. U.S.A.* 102: 5773-5778.

Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* 38: 317-331.

Russell, A.G., Charette, J.M., Spencer, D.F. and Gray, M.W. (2006) An early evolutionary origin for the minor spliceosome. *Nature* 443: 863-866.

Rymond, B.C. and Rosbash, M. (1985) Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. *Nature* 317: 735-737.

Salgado-Garrido, J., Bragado-Nilsson, E., Kandels-Lewis, S. and Séraphin, B. (1999) Sm and Sm-like proteins assemble in two related complexes of deep evolutionary origin. *EMBO J.* 18: 3451-3462.

Saltzman, A.L., Kim, Y.K., Pan, Q., Fagnani, M.M., Maquat, L.E. and Blencowe, B.J. (2008) Regulation of multiple core spliceosomal proteins by alternative splicing-coupled nonsense-mediated mRNA decay. *Mol. Cell. Biol.* 28: 4320-4330.

Sawa, H. and Abelson, J. (1992) Evidence for a base-pairing interaction between U6 small nuclear RNA and the 5' splice site during the splicing reaction in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 89: 11269-11273.

Scamborova, P., Wong, A. and Steitz, J.A. (2004) An intronic enhancer regulates splicing of the twintron of Drosophila melanogaster prospero pre-mRNA by two different spliceosomes. *Mol. Cell. Biol.* 24: 1855-1869.

Schaffert, N., Hossbach, M., Heintzmann, R., Achsel, T. and Lührmann, R. (2004) RNAi knockdown of hPrp31 leads to an accumulation of U4/U6 di-snRNPs in Cajal bodies. *EMBO J.* 23: 3000-3009.

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E. and Zipursky, S.L. (2000) Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell* 101: 671-684.

Schneider, C., Will, C.L., Makarova, O.V., Makarov, E.M. and Lührmann, R. (2002) Human U4/U6.U5 and U4atac/U6atac.U5 tri-snRNPs exhibit similar protein compositions. *Mol. Cell. Biol.* 22: 3219-3229.

Schneider, C., Will, C.L., Brosius, J., Frilander, M.J. and Lührmann, R. (2004) Identification of an evolutionarily divergent U11 small nuclear ribonucleoprotein particle in Drosophila. *Proc. Natl. Acad. Sci. U.S.A.* 101: 9584-9589.

Schor, I.E., Rascovan, N., Pelisch, F., Alló, M. and Kornblihtt, A.R. (2009) Neuronal cell depolarization induces intragenic chromatin modifications affecting NCAM alternative splicing. *Proc. Natl. Acad. Sci. U.S.A.* 106: 4325-4330.

Séraphin, B. (1995) Sm and Sm-like proteins belong to a large family: identification of proteins of the U6 as well as the U1, U2, U4 and U5 snRNPs. *EMBO J.* 14: 2089-2098.

Sharp, P.A. (1985) On the origin of RNA splicing and introns. *Cell* 42: 397-400.

Sharp, P.A. (1991) "Five easy pieces". *Science* 254: 663.

Sharp, P.A. and Burge, C.B. (1997) Classification of introns: U2 type or U12 type. *Cell* 91: 875-879.

Sheth, N., Roca, X., Hastings, M.L., Roeder, T., Krainer, A.R. and Sachidanandam, R. (2006) Comprehensive splice-site analysis using comparative genomics. *Nucleic Acids Res.* 34: 3955-3967.

Shimba, S. and Reddy, R. (1994) Purification of human U6 small nuclear RNA capping enzyme. Evidence for a common capping enzyme for gamma-monomethyl-capped small RNAs. *J. Biol. Chem.* 269: 12419-12423.

Shukla, G.C. and Padgett, R.A. (2001) The intramolecular stem-loop structure of U6 snRNA can functionally replace the U6atac snRNA stem-loop. *RNA* 7: 94-105.

Shukla, G.C. and Padgett, R.A. (2002) A catalytically active group II intron domain 5 can function in the U12-dependent spliceosome. *Mol. Cell* 9: 1145-1150.

Shukla, G.C. and Padgett, R.A. (2004) U4 small nuclear RNA can function in both the major and minor spliceosomes. *Proc. Natl. Acad. Sci. U.S.A.* 101: 93-98.

Singh, J. and Padgett, R.A. (2009) Rates of in situ transcription and splicing in large human genes. *Nat. Struct. Mol. Biol.* 16: 1128-1133.

Sleeman, J.E. and Lamond, A.I. (1999) Newly assembled snRNPs associate with coiled bodies before speckles, suggesting a nuclear snRNP maturation pathway. *Curr. Biol.* 9: 1065-1074.

Smith, D.J., Query, C.C. and Konarska, M.M. (2008) "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol. Cell* 30: 657-666.

Sontheimer, E.J. and Steitz, J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* 262: 1989-1996.

Sorek, R. and Ast, G. (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* 13: 1631-1637.

de Souza, S.J., Long, M., Schoenbach, L., Roy, S.W. and Gilbert, W. (1996) Intron positions correlate with module boundaries in ancient proteins. *Proc. Natl. Acad. Sci. U.S.A.* 93: 14632-14636.

Spies, N., Nielsen, C.B., Padgett, R.A. and Burge, C.B. (2009) Biased chromatin signatures around polyadenylation sites and exons. *Mol. Cell* 36: 245-254.

Spiller, M.P., Boon, K.L., Reijns, M.A. and Beggs, J.D. (2007) The Lsm2-8 complex determines nuclear localization of the spliceosomal U6 snRNA. *Nucleic Acids Res.* 35: 923-929.

Spradling, A.C., Stern, D., Beaton, A., Rhem, E.J., Laverty, T., Mozden, N., Misra, S. and Rubin, G.M. (1999) The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. *Genetics* 153: 135-177.

Staley, J.P. and Guthrie, C. (1998) Mechanical devices of the spliceosome: Motors, clocks, springs, and things. *Cell* 92: 315-326.

Staley, J.P. and Guthrie, C. (1999) An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol. Cell* 3: 55-64.

Stanek, D., Rader, S.D., Klingauf, M. and Neugebauer, K.M. (2003) Targeting of U4/U6 small nuclear RNP assembly factor SART3/p110 to Cajal bodies. *J. Cell Biol.* 160: 505-516.

Stanek, D., Pridalová-Hnilicová, J., Novotný, I., Huranová, M., Blazíková, M., Wen, X., Sapra, A.K. and Neugebauer, K.M. (2008) Spliceosomal small nuclear ribonucleoprotein particles repeatedly cycle through Cajal bodies. *Mol. Biol. Cell* 19: 2534-2543.

Stevens, S.W., Ryan, D.E., Ge, H.Y., Moore, R.E., Young, M.K., Lee, T.D. and Abelson, J. (2002) Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol. Cell* 9: 31-44.

Stoltzfus, A., Spencer, D.F., Zuker, M., Logsdon, J.M.J. and Doolittle, W.F. (1994) Testing the exon theory of genes: the evidence from protein structure. *Science* 265: 202-207. **Sverdlov, A.V., Rogozin, I.B., Babenko, V.N. and Koonin, E.V.** (2003) Evidence of splice signal migration from exon to intron during intron evolution. *Curr. Biol.* 13: 2170-2174.

Sverdlov, A.V., Babenko, V.N., Rogozin, I.B. and Koonin, E.V. (2004) Preferential loss and gain of introns in 3' portions of genes suggests a reverse-transcription mechanism of intron insertion. *Gene* 338: 85-91.

Talerico, M. and Berget, S.M. (1990) Effect of 5' splice site mutations on splicing of the preceding intron. *Mol. Cell. Biol.* 10: 6299-6305.

Talerico, M. and Berget, S.M. (1994) Intron definition in splicing of small Drosophila introns. *Mol. Cell. Biol.* 14: 3434-3445.

Tarn, W. and Steitz, J.A. (1996a) Highly divergent U4 and U6 small nuclear RNAs required for splicing rare AT-AC introns. *Science* 273: 1824-1832.

Tarn, W. and Steitz, J.A. (1996b) A novel spliceosome containing U11, U12 and U5 snRNPs excises a minor class (AT-AC) intron in vitro. *Cell* 84: 801-811.

Tarn, W., Yario, T.A. and Steitz, J.A. (1995) U12 snRNA in vertebrates: Evolutionary conservation of 5' sequences implicated in splicing of pre-mRNAs containing a minor class of introns. *RNA* 1: 644-656.

Tarn, W. and Steitz, J.A. (1997) Pre-mRNA splicing: the discovery of a new spliceosome doubles the challenge. *Trends Biochem. Sci.* 22: 132-137.

Teigelkamp, S., Mundt, c., Achsel, T., Will, C.L. and Lührmann, R. (1997) The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *RNA* 3: 1313-1326.

Tidow, H., Andreeva, A., Rutherford, T.J. and Fersht, A.R. (2009) Solution structure of the U11-48K CHHC zinc-finger domain that specifically binds the 5' splice site of U12-type introns. *Structure* 17: 294-302.

Tonegawa, S., Maxam, A.M., Tizard, R., Bernard, O. and Gilbert, W. (1978) Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci. U.S.A.* 75: 1485-1489.

Toor, N., Keating, K.S., Taylor, S.D. and Pyle, A.M. (2008) Crystal structure of a self-spliced group II intron. *Science* 320: 77-82.

Tseng, C. and Cheng, S. (2008) Both catalytic steps of nuclear pre-mRNA splicing are reversible. *Science* 320: 1782-1784.

Turunen, J.J., Will, C.L., Grote, M., Lührmann, R. and Frilander, M.J. (2008) The U11-48K protein contacts the 5' splice site of U12-type introns and the U11-59K protein. *Mol. Cell. Biol.* 28: 3548-3560.

Tycowski, K.T., You, Z.H., Graham, P.J. and Steitz, J.A. (1998) Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell* 2: 629-638.

Valadkhan, S. and Manley, J.L. (2001) Splicing-related catalysis by protein-free snRNAs. *Nature* 413: 701-707.

Valadkhan, S. (2007) The spliceosome: a ribozyme at heart?. *Biol. Chem.* 388: 693-697.

Valadkhan, S., Mohammadi, A., Jaladat, Y. and Geisler, S. (2009) Protein-free small nuclear RNAs catalyze a two-step splicing reaction. *Proc. Natl. Acad. Sci. U.S.A.* 106: 11901-11906.

Valcárcel, J., Gaur, R.K., Singh, R. and Green, M.R. (1996) Interaction of U2AF65 RS region with pre-mRNA of branch point and promotion base pairing with U2 snRNA. *Science* 273: 1706-1709.

Valcárcel, J. and Gebauer, F. (1997) Post-transcriptional regulation: the dawn of PTB. *Curr. Biol.* 7: R705-8.

Valencia, P., Dias, A.P. and Reed, R. (2008) Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* 105: 3386-3391.

Van Aken, O., Pecenková, T., van de Cotte, B., De Rycke, R., Eeckhout, D., Fromm, H., De Jaeger, G., Witters, E., Beemster, G.T.S., Inzé, D. and Van Breusegem, F. (2007) Mitochondrial type-I prohibitins of Arabidopsis thaliana are required for supporting proficient meristem development. *Plant J.* 52: 850-864.

Verbeeren, J., Niemelä, E.H., Turunen, J.J., Will, C.L., Ravantti, J.J., Lührmann, R. and Frilander, M.J. (2010) An ancient mechanism for splicing control: U11 snRNP as an activator of alternative splicing. *Mol. Cell* 37: 821-833.

Vibranovski, M.D., Sakabe, N.J., de Oliveira, R.S. and de Souza, S.J. (2005) Signs of ancient and modern exon-shuffling are correlated to the distribution of ancient and modern domains along proteins. *J. Mol. Evol.* 61: 341-350.

Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T., Bird, A.C., Hunt, D.M. and Bhattacharya, S.S. (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell* 8: 375-381.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P. and Burge, C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456: 470-476.

Wang, G. and Cooper, T.A. (2007) Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat. Rev. Genet.* 8: 749-761.

Wang, H., Gao, M.X., Li, L., Wang, B., Hori, N. and Sato, K. (2007) Isolation, expression, and characterization of the human ZCRB1 gene mapped to 12q12. *Genomics* 89: 59-69.

Wang, Z. and Burge, C.B. (2008) Splicing regulation: From a parts list of regulatory elements to an integrated splicing code. *RNA* 14: 802-813.

Wassarman, D.A. and Steitz, J.A. (1992a) Interaction of small nuclear RNA's with precursor messenger RNA during in vitro splicing. *Science* 257: 1918-1925.

Wassarman, K.M. and Steitz, J.A. (1992b) The low-abundance U11 and U12 small nuclear ribonucleoproteins (snRNPs) interact to form a two-snRNP complex. *Mol. Cell. Biol.* 12: 1276-1285.

Wassarman, K.M. and Steitz, J.A. (1993) Association with terminal exons in pre-mRNAs: a new role for the U1 snRNP?. *Genes Dev.* 7: 647-659.

Will, C.L., Schneider, C., Reed, R. and Lührmann, R. (1999) Identification of both shared and distinct proteins in the major and minor spliceosomes. *Science* 284: 2003-2005.

Will, C.L. and Lührmann, R. (2001) Spliceosomal UsnRNP biogenesis, structure and function. *Curr. Opin. Cell Biol.* 13: 290-301.

Will, C.L., Schneider, C., Hossbach, M., Urlaub, H., Rauhut, R., Elbashir, S., Tuschl, T. and Lührmann, R. (2004) The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *RNA* 10: 929-941.

Wu, Q. and Krainer, A.R. (1996) U1-mediated exon definition interactions between AT-AC and GT-AG introns. *Science* 274: 1005-1008.

Wu, Q. and Krainer, A.R. (1997) Splicing of a divergent subclass of AT-AC introns requires the major class spliceosomal snRNAs. *RNA* 3: 586-601.

Wu, Q. and Krainer, A.R. (1998) Purine-rich enhancers function in the AT-AC pre-mRNA splicing pathway and do so independently of intact U1 snRNP. *RNA* 4: 1664-1673.

Wu, Q. and Krainer, A.R. (1999) AT-AC pre-mRNA splicing mechanism and conservation of minor introns in voltage-gated ion channel genes. *Mol. Cell. Biol.* 19: 3225-3236.

Yean, S.L., Wuenschell, G., Termini, J. and Lin, R.J. (2000) Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature* 408: 881-884.

Yeo, G., Holste, D., Kreiman, G. and Burge, C.B. (2004) Variation in alternative splicing across human tissues. *Genome Biol.* 5: R74.

Yu, Y. and Steitz, J.A. (1997) Site-specific crosslinking of mammalian U11 and U6atac to the 5' splice site of an AT-AC intron. *Proc. Natl. Acad. Sci. U.S.A.* 94: 6030-6035.

Yu, Y., Scharl, E.C., Smith, C.M. and Steitz, J.A. The growing world of small nuclear ribonucleoproteins in *The RNA world* (eds. Gesteland, R.F., Cech, T.R. and Atkins, J.F.) 2nd edition, pp. 487-524 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1999).

Yu, Y., Shu, M., Narayanan, A., Terns, R.M., Terns, M.P. and Steitz, J.A. (2001) Internal modification of U2 small nuclear (sn)RNA occurs in nucleoli of Xenopus oocytes. *J. Cell Biol.* 152: 1279-1288.

Zahler, A.M., Lane, W.S., Stolk, J.A. and Roth, M.B. (1992) SR proteins - a conserved family of pre-messenger-RNA splicing factors. *Genes Dev.* 6: 837-847.

Zamore, P.D. and Green, M.R. (1989) Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc. Natl. Acad. Sci. U.S.A.* 86: 9243-9247.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M. and Dreyfuss, G. (2008) SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. *Cell* 133: 585-600.

Zhao, E., Li, J., Xie, Y., Jin, W., Zhang, Z., Chen, J., Zeng, L., Yin, G., Qian, J., Wu, H., Ying, K., Zhao, R.C. and Mao, Y. (2003) Cloning and identification of a novel human RNPC3 gene that encodes a protein with two RRM domains and is expressed in the cell nucleus. *Biochem. Genet.* 41: 315-323.

Zhou, Z., Licklider, L.J., Gygi, S.P. and Reed, R. (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* 419: 182-185.

Zhu, W. and Brendel, V. (2003) Identification, characterization and molecular phylogeny of U12-dependent introns in the Arabidopsis thaliana genome. *Nucleic Acids Res.* 31: 4561-4572.

Zorio, D.A.R. and Blumenthal, T. (1999) Both subunits of U2AF recognize the 3' splice site in Caenorhabditis elegans. *Nature* 402: 835-838.